Biochemical and Structural Studies of Dimethylarginine Dimethylaminohydrolase-1, the ABC Transporter Superfamily and the Sodium Citrate Transporter CitS.

DISSERTATION

ZUR

ERLANGUNG DER NATURWISSENSCHAFTLICHEN DOKTORWÜRDE
(DR. SC. NAT.)

VORGELEGT DER

MATHEMATISCHEN-NATURWISSENSCHAFTLICHEN FAKULTÄT

DER

UNIVERSITÄT ZÜRICH

VON

DANIEL FREY

AUS

DEUTSCHLAND

PROMOTIONSKOMITEE

PROF. DR. MARKUS G. GRÜTTER (VORSITZ)

PROF. DR. MILAN VAŠÁK

PROF. DR. RAIMUND DUTZLER

ZÜRICH 2007

To my family

Table of contents

Summary **1**

Zusammenfassung **4**

Chapter A - DDAH ***7***

A.1 Introduction to Dimethylarginine Dimethylaminohydrolase **7**

A.1.1 Nitric Oxide 7

A.1.2 Family of Arginine Modifying Enzymes 12

A.2 Dimethylarginine Dimethylaminohydrolase I **14**

A.3 Structure of the Mammalian NOS Regulator Dimethylarginine Dimethylaminohydrolase (DDAH-1): A Basis for the Design of Specific Inhibitors **15**

Chapter B - Membrane Proteins ***27***

B.1 Introduction to Membrane Proteins **27**

B.2 Membrane Protein Classification **28**

B.3 Transport Proteins **30**

B.4 Detergents **33**

B.5 Crystallization of Membrane Proteins **36**

Chapter C - ABC Transporters ***41***

C.1 Introduction to ABC Transporters **41**

C.1.1 Sequence Similarity 43

C.1.2 Domain Arrangement of ABC Transporters 45

C.1.3 Structural Models 47

C.1.4 Transport Energy Proliferation and ATP-switch Model 52

C.2 Results for ABC Transporters **54**

C.2.1 Eukaryotic ABC Transporter 54

C.2.1.1 Protein Expression in Insect Cells 55

C.2.1.2 Solubilization of mBsep 56

C.2.1.3 Cloning and Expression of mBsep in *Escherichia Coli* 57

C.2.1.4 Cloning and Expression of mBsep in Mammalian Cells 58

C.2.2 Prokaryotic ABC Transporters 60

C.2.2.1 Strategy 60

C.2.2.2 Selection 62

C.2.2.3 Cloning of Prokaryotic ABC Transporters 65

C.2.2.4 Expression of Prokaryotic ABC Transporters 67

C.2.2.5 Purification 68

C.2.3 Protocols for Specific Prokaryotic ABC Transporters 70

C.2.3.1 Proteins Originating from *Thermotoga maritima* 70

C.2.3.2 Proteins Originating from *Escherichia Coli* 80

C.2.3.3 Crystallization of ABC transporters 95

C.3 Discussion and Conclusion **96**

Chapter D - CitS 101

D.1 Introduction to the Sodium Citrate Symporter CitS 101

D.1.1 Anaerobic Citrate Metabolism: Citrate Fermentation 102

D.1.2 The Sodium Citrate Symporter CitS of *Klebsiella Pneumoniae* 104

D.2 Results of the Sodium Citrate Symporter CitS 107

D.2.1 Improvement of Crystal Growth 107

D.2.2 Post Crystallization Improvement of CitS Crystals 112

D.2.3 Crystallographic Analysis 115

D.2.4 Heavy Atom Derivative Preparation of CitS Crystals 118

D.2.4.1 Isomorphous Derivatives using Heavy Metals 119

D.2.4.2 Derivatization with L-selenomethionine 121

D.3 Crystallization of CitS - Complexes 122

D.3.1 Complex Crystallization of CitS with Fab Fragments 122

D.3.2 Complexes of CitS with selected DARPin Proteins 122

D.3.2.1 Crystallization of the CitS-DARPin Complexes 123

D.3.2.2 Crystallographic Analysis of CitS-DARPin Complexes 125

D.4 Discussion and Conclusion 128

Chapter E - Structure of the Fab Fragment f3p4 132

E.1 Unusual twinning in crystals of the CitS binding antibody Fab fragment f3p4 132

Chapter F - Appendices 161

F.1 Crystallographic Methods 161

F.2 Biochemical Methods 169

F.3 ABC Transporters 170

F.4 CitS 181

F.5 β-Catenin 184

Chapter G - References 190

Chapter H 201

Acknowledgement 201

Curriculum vitae 203

Publications and Meeting contributions 204

Summary

The area of structural biology covers a broad range of different techniques and methods that are used to obtain valuable information on the protein of interest. This thesis describes the expression and purification properties as well as biochemical and functional aspects of several proteins. The first section provides insight in the crystal structure of the protein Dimethylarginine Dimethylaminohydrolase-1 (DDAH-1) of bovine brain. The second part of the thesis covers the work on membrane proteins including the ABC transporter superfamily and the protein CitS, a member of the 2-hydroxycarboxylate transporter family.

DDAH metabolizes methylated arginine derivatives, which are naturally occurring inhibitors of all three isoforms of nitric oxide synthases (NOS). Since DDAH activity regulates nitric oxide concentration, the protein emerges as promising drug target. Crystal structures of DDAH 1 of bovine brain were solved, including DDAH 1 alone and in complex with Zn^{2+}, L- citrulline, L-homocysteine or S-nitroso L-homocysteine. The later forms a covalent bond with the active site cysteine and, therefore, is an endogenous suicide inhibitor. The high-resolution structures with resolutions up to 1.08 Å provided valuable insights into (i) the structure of the binding pocket, (ii) the active site and a flexible loop region that is important for the binding of the substrate, and (iii) the coordination sphere of the Zn^{2+} ion that is exclusively present in eukaryotic DDAH at high and low pH values. Homology modeling of the isoenzyme DDAH-2 revealed differences in the binding pocket of the two enzymes. These variations might prove to be useful for the development of isoform specific drugs against DDAH leading to tissue specific regulation of nitric oxide synthases.

The second part of the thesis covers the work that was done in the field of membrane proteins. Two classes of transport systems, ABC transporters and 2-hydroxycarboxylate transporters were analyzed, where the work on ABC transporters mainly dealt with the expression, solubilization and purification of these proteins. Present in all genomes, starting from simple archea to bacteria and complex eukaryotes, ABC transporters enclose a huge superfamily of membrane proteins with more than 2000 family members. The versatile functions include the regulation of cellular processes, the import of nutrients and the export of peptides or toxic substances. The family members share a common mechanistic principle: The chemical energy, stored in the phosphate ester bond of the ATP molecule, is used to facilitate the transport of the substrate across the membrane. ABC transporters of various origins were selected to enhance the chance of obtaining the large quantities of purified protein required for crystallographic studies. Analysis of expression, solubilization and purification efficiencies resulted in three different transport systems that were analyzed in further biochemical experiments. ATP binding or ATP hydrolysis experiments in conjunction with FT-IR confirmed the presence of folded material in the detergent micelle. The oligomeric state of the proteins in the presence of different detergents was assessed by size exclusion chromatography studies either with or without a subsequent static light scattering analysis. Initial crystallization experiments with the purified ABC transporter were conducted.

The second project in the field of membrane proteins covers the work on the protein CitS and in this thesis the focus lies on the crystallization of the protein and the analysis of the recorded X-ray data. The transporter CitS belongs to the 2-hydroxycarboxylate-transporter family and facilitates the simultaneous uptake of sodium ions and citrate, enabling the bacterium *Klebsiella Pneumoniae* to use citrate as sole carbon and energy source. Characterization of the purified protein by size exclusion chromatography and static light scattering revealed a dimeric assembly of the protein in solution. The protein crystallized in space group $P2_1$ and the crystals diffracted up to 3.6 Å. The self-rotation function of the native data set showed presumably 2 dimeric proteins that assemble in a tetramer in the crystals. No homologues of the protein are characterized structurally so far and therefore phases have to be determined experimentally in order to solve the structure. Multiple methods were considered to obtain the phase information: (i) soaking or co-crystallization of the crystals with heavy atoms, and (ii) co-crystallization of the membrane protein with artificial binding partners to obtain initial phase information of the complex by molecular replacement. In an effort to co-crystallize Fab fragments or designed ankyrin repeat proteins with the membrane protein, crystals of a complex of CitS with an ankyrin were obtained. These crystals diffract up to 4 Å and belong to space group P1 containing presumably 8 molecules in the unit cell. The tetramer present in crystals of CitS in space group $P2_1$ is absent in the crystals of the complex. The binding molecules alone were characterized structurally and the structure solution and refinement of the twinned crystals of the Fab fragment f3p4 at 1.9 Å resolution is reported. Determination of the phases and the structure solution of the membrane protein CitS, either alone or in complex with the artificial binding proteins is in progress.

Zusammenfassung

Die Strukturbiologie umfaßt eine Vielzahl verschiedener Techniken und Methoden, mit denen wertvolle Informationen über Proteine bestimmt werden können. Die vorliegende Doktorarbeit beschreibt die Herstellung und Reinigung sowie die biochemischen und funktionellen Eigenschaften mehrerer Proteine. Die Arbeit ist in zwei Teile gegliedert, wobei der erste Abschnitt die Röntgenstrukturanalyse des Proteins Dimethylarginin Dimethylaminohydrolase-1 (DDAH-1) beinhaltet. Der zweite Abschnitt der Arbeit befaßt sich mit Membranproteinen, im Speziellen mit ABC Transportern und CitS, einem Mitglied der 2-Hydroxycarboxylat Transporter Familie.

Methylierte Argininderivate sind natürlich vorkommende Inhibitoren der drei Isoformen der Stickstoffmonoxidsynthase (*nitric oxide synthase*), welche von DDAH abgebaut werden. Da DDAH die Stickstoffmonoxid Konzentration beeinflußt, ist das Protein ein vielversprechendes Zielmolekül für die Entwicklung neuer Medikamente. Die Kristallstruktur der DDAH-1 aus Rinderhirn wurde als freies Enzym und im Komplex mit Zn^{2+}, L-Citrullin, L-Homocystein oder S-Nitroso L-Homocystein gelöst. S-Nitroso L-Homocystein ist ein Suizid Inhibitor der DDAH, da es eine kovalente Bindung mit dem Cystein des Reaktionszentrums eingeht. Die hochaufgelösten Strukturen mit einer Auflösung von bis zu 1.08 Å gewähren einen detaillierten Einblick in die Funktionsweise des Enzyms. Unsere Arbeiten gaben Aufschluß über: (i) die Bindungstasche, (ii) das Reaktionszentrum und einem beweglichen Teil des Proteins, der für die Substratbindung von Bedeutung ist, und (iii) die Koordination des Zn^{2+} Ions welche bei verschiedenen pH Werten bestimmt werden konnte. Ein Modell des Isoenzyms DDAH 2 zeigt die Unterschiede in den Bindungstaschen der beiden Enzyme auf. Diese Unterschiede könnten sich bei der Entwicklung spezifischer Inhibitoren gegen die verschiedenen Isoformen der DDAH als nützlich erweisen, welche eine zelltyp-abhängige Regulation der Stickstoffmonoxidsynthasen erlauben würden.

Im zweiten Teil der Arbeit werden die Experimente beschrieben, die an Membranproteinen ausgeführt wurden. Zwei Klassen von Membran Transportproteinen wurden bearbeitet: ABC Transporter aus verschiedenen Organismen und der 2-Hydroxycarboxylat Transporter aus *Klebsiella Pneumoniae*. Die ABC Transporter wurden im Bezug der Expression, Solubilisierung und Aufreinigung biochemisch untersucht. ABC Transporter gibt es in allen Genomen, von einfachen Archaebakterien, über Bakterien bis zu komplexen Eukaryonten. Sie bilden eine größten Familien unter den Membranproteinen mit mehr als 2000 Proteinen. Ihre vielfältigen Funktionen, darunter die Regulation von Zellprozessen, der Import von Nährstoffen und der Export von Peptiden oder toxischen Substanzen haben ein gemeinsames mechanistisches Prinzip; die chemische Energie, welche in der Phosphoesterbindung des ATP Moleküls gespeichert ist, wird verwendet um den Transport der Substrate durch die Membran zu ermöglichen. Eine Auswahl von ABC Transportern von verschiedenen Organismen wurde getroffen, um die Chance zu erhöhen, große Mengen an aufgereinigtem Protein zu erhalten, welche für kristallographische Experimente benötigt werden. Nach Expressions-, Löslichkeits- und Aufreinigungsexperimenten verblieben drei verschiedene ABC Transportsysteme, welche anschließend detaillierter biochemisch analysiert werden konnten. ATP bindende oder hydrolysierende Experimente in Verbindung mit FT-IR bestätigten, daß gefaltetes Protein in den Detergenzmizellen vorhanden war. Der oligomere Zustand des Proteins wurde in Gegenwart von verschiedenen Detergenzien durch Gelfiltration, entweder mit oder ohne gekoppelter Statische Lichtstreuungs Analyse ermittelt. Kristallisationsexperimente mit den aufgereinigten ABC Transportern wurden durchgeführt.

Die Arbeiten mit dem Membranprotein CitS aus *Klebsiella Pneumoniae* betrafen hauptsächlich die Kristallisation und die Analyse der mit Röntgendiffraktionsdaten. Der Transporter CitS gehört zur 2-Hydroxycarboxylat Familie und ermöglicht den parallelen Import von Natrium Ionen und Citrat, was dem Bakterium *Klebsiella pneumoniae* erlaubt mit Citrat als einzige Kohlenstoff- und Energiequelle zu überleben. Das Protein wurde in *Escherichia Coli* exprimiert, in Detergenz solubilisiert, und anschließend aufgereinigt. Die nachfolgende Charakterisierung mittels Gelfiltration und statischer Lichtstreuung ließ erkennen, daß das Protein in Lösung als Dimer vorliegt. Das Protein kristallisierte in der Raumgruppe $P2_1$. Die Kristalle streuten bis 3.6 Å. Die Eigenrotationsfunktion des nativen Datensatzes wies auf 2 dimere Proteine hin, welche sich im Kristall zu einem Tetramer zusammenlagerten. Homologe des Proteins wurden strukturell bis jetzt nicht charakterisiert, daher müssen für die Röntgenstruktur die Phasen experimentell ermittelt werden. Verschiedene Methoden die Phaseninformation zu erhalten wurden angewandt: (i) Inkubation oder Co-Kristallisation mit Schwermetallen, (ii) Co-Kristallisation des Membranproteins mit einem künstlich hergestellten Bindungspartner welche es ermöglichen, erste Phaseninformationen zu erhalten und die Struktur danach durch Molekularen Ersatz zu lösen. CitS wurde im Komplex mit den Fab-Fragmenten f3p4 und fm4p3 und den Ankyrin Proteinen cp34h_15 und Na1F kristallisiert. Kristalle von CitS im Komplex mit dem Ankyrin cp34h_15 konnten gezüchtet werden und streuten bis 4 Å. Sie gehören zur Raumgruppe P1 und haben wahrscheinlich 8 Komplexmoleküle in der Einheitszelle. Ein Tetramer, wie in den CitS-Kristallen der Raumgruppe $P2_1$ gefunden, konnte hier nicht beobachtet werden. Die Bindungspartner f3p4, cp34h_15 und Na1F ohne CitS wurden strukturell charakterisiert und die Strukturaufklärung des verzwillingten Kristalls des Fab-Fragmentes f3p4, welcher bis 1.9 Å streute, werden hier erläutert. Die Bestimmung der Phasen und die strukturelle Aufklärung des Membranproteins CitS, mit und ohne Bindungspartner, sind noch in Arbeit.

Chapter A - DDAH

The protein dimethylarginine dimethylaminohydrolase-1 (Dimethylarginase-1, DDAH-1) hydrolyzes methyl derivatives of arginine, molecules that are known to inhibit nitric oxide synthases. The relationship between the two enzymes enables an indirect regulation of nitric oxide levels in the cells. Active dimethylarginine dimethylaminohydrolase promotes a higher activity of nitric oxide synthases and therefore leads to higher nitric oxide levels. This work describes various crystallographic structures of the isoform I of bovine dimethylarginine dimethylaminohydrolase.

A.1 Introduction to Dimethylarginine Dimethylaminohydrolase

A.1.1 Nitric Oxide

Research in the last decade established the gaseous free radical nitric oxide (NO) as a molecule involved in signaling of different important biological processes. It is a major signaling molecule in neurons and in primary defense mechanisms against microorganisms in macrophages. Furthermore NO is involved in the endothelium dependent regulation of blood flow and blood pressure as well as in the inhibition of the activation of blood platelets (Bruckdorfer, 2005).

NO is synthesized enzymatically from the amino acid L-arginine in a number of tissues by different isoforms of the enzyme nitric oxide synthase (NOS). The different isoforms of NOS have been well characterized and three isoforms are predominant: inducible NOS (iNOS/NOS II), endothelial NOS (eNOS/NOS III) and neuronal NOS (nNOS/NOS I). Although these names are frequently used, they are misleading since there is no strict division of the isoforms and cells often express more than one isoenzyme. iNOS produces large amounts of NO during immune and inflammatory responses and is induced transcriptionally. NF-κB, which is activated by cytokines and lipopolysaccharides, induces the expression of the gene encoding iNOS and, therefore, leads to an elevated level of NO. In contrast, eNOS and nNOS are constitutively expressed and are activated by calmodulin (CaM) and calcium. nNOS and eNOS produce NO at a lower levels than iNOS. They are involved in hormonal and neuronal signaling as well as in the regulation of blood flow and blood pressure.

The human isoforms have a molecular weight between 130 and 160 kDa and contain a central CaM binding site, which separates the N-terminal oxygenase domains and the C-terminal reductase domain (Figure 1).

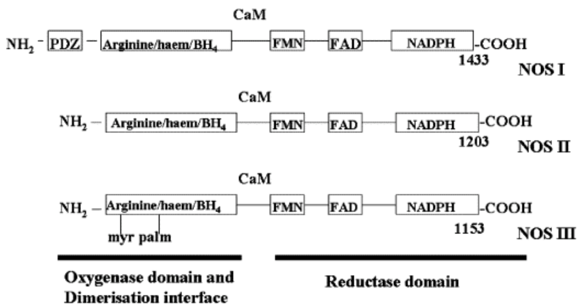

Figure 1: General structure overview of the three isoform of NOS. NOS I (nNOS), NOS II (iNOS) and NOS III (eNOS) with their oxidation N-terminal domains and the reductive C-terminal domains. Between the two domains is the binding site for calmodulin that activates the NOS.

The active enzymes are homodimers with a Zn^{2+} atom bound at the interface of their N-terminal domains (Figure 2). Although the specific process of NO production includes various oxido/reductive steps, requiring a large number of cofactors bound to specific binding sites in the enzymes, the production of NO and conversion of L-arginine to L-citrulline with the reaction intermediate N-hydroxy-L-arginine can be summarized by the reaction

$$\text{L-arginine} + 2\ \text{NADPH} + O_2 \rightarrow \text{L-citrulline} + \text{NO} + 2\ \text{NADP}^+$$

The N-terminal oxygenase domain contains the binding sites of the two cofactors heme–Fe and the tetrahydrobiopterin as well as the binding pocket of the substrate L-arginine (Figure 2), whereas the C-terminal reductase domain binds flavin mononucleotide (FMN), flavin adenine dinucleotide (FAD) and nicotine adenine dinucleotide phosphate (NADPH). The connecting CaM domain regulates the electron flow between the reductase and the oxygenase domain and therefore regulates the production of NO.

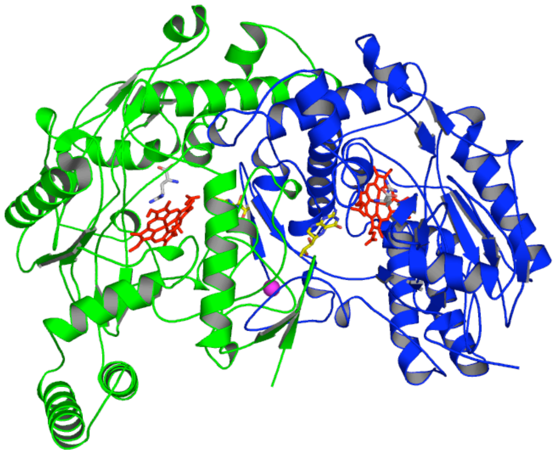

Figure 2: Oxygenase domain of iNOS (PDB ID: 1NLI). For enzymatic activity, nitric oxide synthase (NOS) enzymes have to form a homodimer. The dimer interface locates a Zn^{2+} atom, which is necessary for full activity (magenta). The substrate arginine, the co-factors tetrahydrobiopterin and the heme group are shown in white, yellow and red, respectively.

The half-life of only a few seconds illustrates the high reactivity of the NO radical *in vivo*. However, since NO is soluble in both aqueous and lipid media, it readily diffuses through the cytoplasm and plasma membranes reacting with the cell itself, or those adjacent where it is synthesized. NO reacts with oxygen and water to produce nitrates and nitrites. NO toxicity is linked to its ability to combine with superoxide anions (O_2^-) to form peroxynitrite ($ONOO^-$), an oxidizing free radical, which reacts with proteins and nucleotides and in addition, enables lipid peroxidation (Szabo 2003). In the mitochondria, peroxynitrite acts on the respiratory chain (I-IV) complex and manganese superoxide dismutase (MnSOD), to generate superoxide anions and hydrogen peroxide, respectively. In macrophages the high toxicity of the NO is used as part of the non specific immune defense mechanism against invading microorganisms (Vallance & Leiper, 2002).

A direct reaction between NO and proteins has been observed in many cases. NO reacts for example with the iron atom of the haem moiety in the active site of the enzyme guanylyl cyclase (GC), activating the enzyme a thousand fold. The resulting increase in levels of the signaling molecule cyclic GMP produced by GC initiates the physiological actions similar to other hormones.

Alternative reactions of NO include direct modification of proteins via S-nitrosylation, nitration of thiol groups and aromatic amino acids. In vivo it has been demonstrated that NO reacts with thiols to form S-nitrosothiols, such as S-nitrosocysteine and S-nitrosoglutathione and it is also present bound to plasma albumin (Al-Sa'doni & Ferro, 2000). Proteins can be activated upon NO modification, e.g. guanylyl cyclase, thioredoxin or prostaglandin H synthase (Hanafy *et al.*, 2001, Haendeler *et al.*, 2002, Goodwin *et al.*, 1998) or deactivated e.g., aconitase, dimethylarginine dimethylaminohydrolase, ribonucleotide reductase or NOS itself (Drapier & Hibbs, 1996, Leiper *et al.*, 2002, Guittet *et al.*, 1998, Mitchell *et al.*, 2005, Gow *et al.*, 2004). Protein modifications by NO are even evident in inflammatory processes and play a role in apoptosis by inhibition of caspase-3 through S-nitrosylation of the active site cysteine (Rossig *et al.*, 1999).

The high toxicity of NO demands strict control and regulation of its levels and of tissues where the radicals are produced. The localization and expression of the different isoforms reflects the various roles of NO. eNOS, involved in the regulation of vascular tone, is docked to the plasma membrane by the hydrophobic anchors myristic and palmitic acid. nNOS contains a PDZ domain that targets the protein to sites in nerve synapses. The two enzymes eNOS and nNOS are activated upon binding of calmodulin, which is triggered by Ca^{2+}. In contrast, Ca^{2+} cannot regulate iNOS, since the affinity of iNOS to CaM is higher and the two proteins form a stable complex in presence or absence of Ca^{2+}. iNOS is transcriptionally regulated and induced in various cells in response to the stimulation by bacteria or cytokines (Vallance & Leiper, 2002). Other factors that influence the bioavailability of NO are the numerous co-factors involved in the synthesis of NO and it is clear that any deficiency will influence the activity of NOS (Vallance & Chan, 2001). Furthermore the activity of NOS may in part be determined by the presence of naturally occurring inhibitors. The substrate analogues of L-arginine, e.g. the methylated arginine derivatives are well known to inhibit all three isoforms of NOS.

Posttranslational modifications of proteins include methylation of arginine side chains. The enzymes protein-arginine-N-methyltransferases (PRMT) transfer methyl groups from S-adenosyl-L-methionine to the guanidinium group of L-arginine. The modifications by PRMT include a transfer of one or two methyl groups to the substrate. Three forms of methylated arginines are known in eukaryotes and they are produced by two classes of PRMT: Type I enzymes produce N^{ω}-monomethylarginine (MMA) and asymmetric N^{ω}, N^{ω}-dimethylarginine (ADMA) whereas proteins belonging to type II also catalyze the formation of MMA and symmetric N^{ω}, $N^{\omega`}$- dimethylarginine (Figure 3) (Bedford & Richard, 2005). Proteolysis of methylated proteins produces the free methylarginines, since these modified residues are not re-incorporated during translation. The free methylarginines MMA and ADMA, but not SDMA, are inhibitors of all three isoforms of NOS. The pool size of arginine derivatives is controlled by renal secretion or metabolism by the enzyme dimethylarginine dimethylaminohydrolase, which hydrolyzes MMA and ADMA, but not SDMA. ADMA is metabolized >90 % by DDAH whereas SDMA is mainly lost in the urine (Tran *et al.*, 2003).

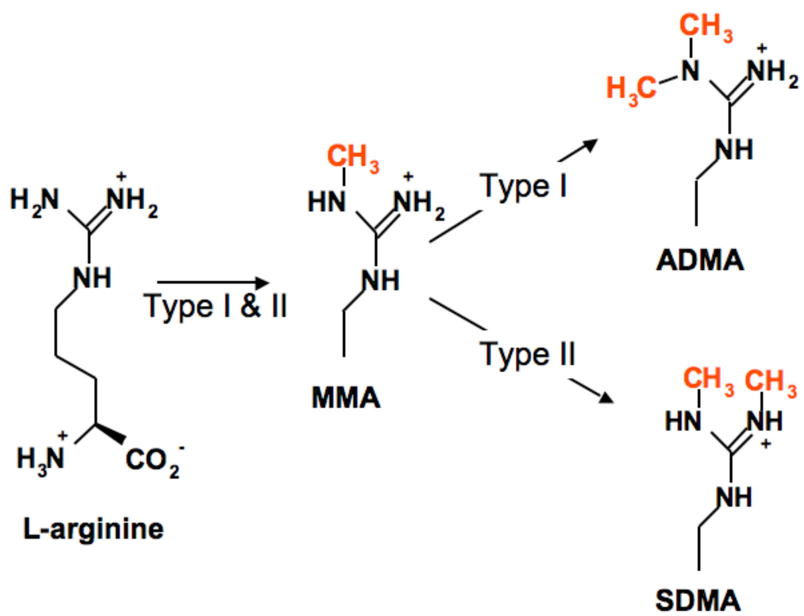

Figure 3: Methylated arginine derivatives. Type I & II PRMTs convert L-arginine to monomethylated L-arginine. Type I enzymes produce asymmetric dimethylated L-arginine whereas symmetric dimethylated L-arginine is produced by type II enzymes.

A.1.2 Family of Arginine Modifying Enzymes

DDAH belongs to the family of arginine-modifying or substituted arginine-modifying enzymes (Shirai *et al.*, 2001, Murray-Rust *et al.*, 2001, Das *et al.*, 2004). The enzymes belonging to this class share low sequence identity, but the overall fold of the active site and the corresponding active site residues are well conserved. The bacterial enzyme arginine deiminase (ADI) catalyzes the deimination of the guanidine group of L-arginine to produce ammonia and L-citrulline, which is the first step of the generation of ATP by the arginine deiminase pathway (Galkin *et al.*, 2004, Das *et al.*, 2004). The C-terminal domain of the Ca^{2+} dependent protein peptidylarginine deiminase (PAD), a protein involved in the conversion of protein arginine residues to citrulline, belongs to this group of enzymes (Arita *et al.*, 2004). PAD catalyzes the deimination of both arginine and MMA but not ADMA and SDMA to citrulline (Bedford & Richard, 2005). Note, that this conversion protects the arginine side chains from the methylation by the enzyme PRMT (Bedford & Richard, 2005). Enzymes that also belong to this class of enzyme are the L-arginine: glycin amidinotransferase (AT) and the enzyme L arginine: inosamine-phosphate amidinotransferase (IPAT). AT is involved in the first step of the creatine biosynthesis (Humm *et al.*, 1997) and IPAT catalyzes reactions involved in the biosynthesis of the streptomycin family of antibiotics (Fritsche *et al.*, 1998). All enzymes of this class share a common fold, which consists of a repetitive ββαβ module arranged in cyclic 5-fold pseudo symmetry (Figure 4).

Although the enzymes bind their substrates in different orientations, (Murray-Rust *et al.*, 2001, Das *et al.*, 2004) they all share a common catalytic mechanism and their catalytic cysteine-histidine-aspartate (glutamate, respectively) triads superimpose well onto each other. Superimposition of the different structures shows that substrate binding is similar in DDAH and ADI. Although the guanidine carbon atoms and the scissile C-N bond overlay in the enzyme AT, the binding of the substrate is different.

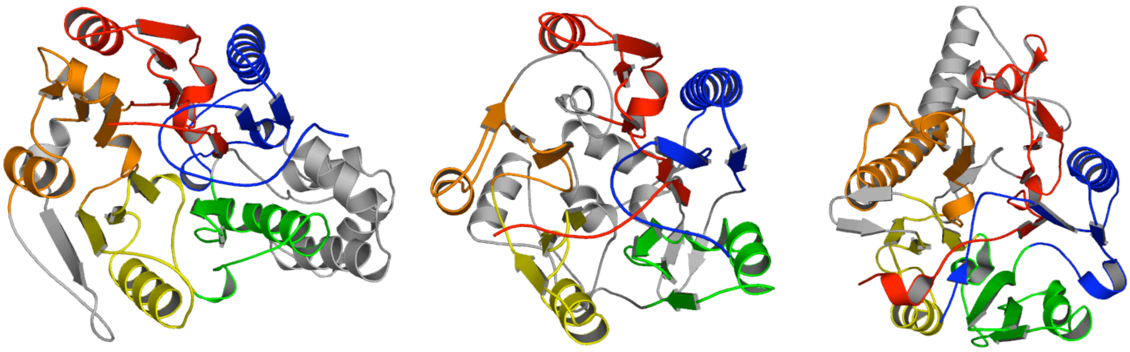

Figure 4: Ribbon representation of members of the arginine or substituted arginine modifying enzyme superfamily. Picture shows ADI, AT and the C-terminal domain of PAD4 from left to the right, respectively (Das *et al.*, 2004, Humm *et al.*, 1997, Arita *et al.*, 2004). The blades are colored from the N to the C-terminus in blue, green, yellow, orange and red, respectively. Gray helices and β-strands are additional insertions in the protein sequences. Note that the last β-strand at the C-terminus interacts with the first three-stranded β-sheet.

The enzymes that modify free arginine or derivatives of free arginine bury their substrates completely in the active site, whereas the substrate of PAD4 is protein bound and projects only partly into the active site of the enzyme. The substrates of the different enzymes ADI, AT, PAD, and DDAH are very similar and the different reaction schemes are shown in Figure 5.

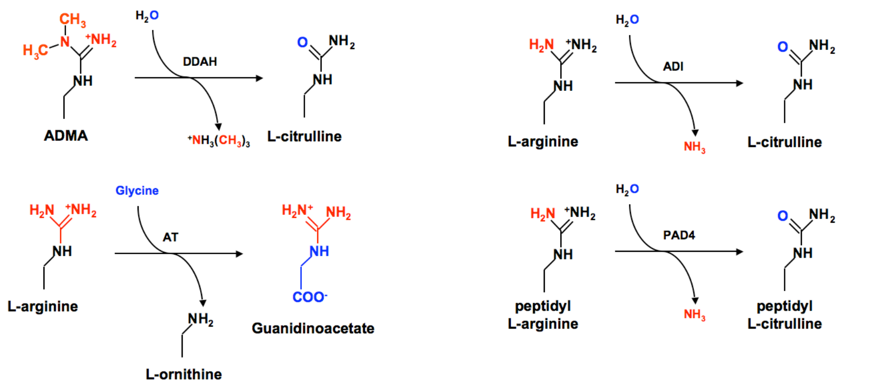

Figure 5: Reactions catalyzed by different enzymes of the arginine or derivatives of arginine modifying enzymes. Although PAD4 utilizes protein bound arginine, the overall reaction compared to ADI is similar. Both transfer an oxygen atom to L-arginine and thereby produce ammonia (two right panels). DDAH catalyzes the reaction of methylated arginine to citrulline and methylamine (upper left image) whereas AT transfers the amidino moiety from arginine to glycine to produce ornithine and guanidinoacetate (lower left scheme).

A.2 Dimethylarginine Dimethylaminohydrolase I

The enzyme L-N^{ω}, N^{ω} -dimethylarginine dimethylaminohydrolase (EC 3.5.3.18) was discovered and characterized by Ogawa *et al* (Ogawa *et al.*, 1989, 1987). The enzyme catalyzes the hydrolysis of ADMA to L-citrulline and dimethylamine or the conversion of MMA to citrulline and monomethylamine (Figure 5).

Activity of the enzyme was found in the kidney, pancreas, liver, and brain with a substrate preference of ADMA over MMA (Ogawa *et al.*, 1987). Further studies on the distribution of DDAH in different tissues as well as the similar substrate specificity supported the idea that DDAH might be a regulator of NOS (Bogumil *et al.*, 1998, Kimoto *et al.*, 1993, Tran *et al.*, 2000). Pharmacological studies on DDAH showed that an inhibition of DDAH leads to an increase in the concentration ADMA. The inhibitor has no direct effect on eNOS but causes endothelium-dependent contraction of aortic rings. This effect was reversed by L-arginine. These findings supported the idea that inhibition of DDAH leads to an accumulation of ADMA and thereby decreases the activity of NOS (MacAllister *et al.*, 1996).

Two isoforms of the enzyme, DDAH I and DDAH II, have been identified in eukaryotes (Tran *et al.*, 2000, Leiper *et al.*, 1999). Both isoforms share a functional homology but they differ in their tissue distribution. DDAH I is mainly expressed in the brain and kidney whereas DDAH II is predominantly present in the heart, spleen, thymus, peripheral leukocytes, lymph node and fetal tissues. The different tissue distribution of the two DDAH isoforms and their correlation with NOS expression pattern led to the assumption that DDAH I might control the activity of iNOS and that DDAH II regulates the ADMA concentration to modulate the activity of eNOS and iNOS (Tran *et al.*, 2000, Bogumil *et al.*, 1998).

A.3 Structure of the Mammalian NOS Regulator Dimethylarginine Dimethylaminohydrolase (DDAH-1): A Basis for the Design of Specific Inhibitors

Daniel Frey, Oliver Braun, Christophe Briand, Milan Vašák
and Markus G. Grütter

Structure 14, 901–911, May 2006 ©2006 Elsevier Ltd All rights reserved DOI 10.1016/j.str.2006.03.006

Structure of the Mammalian NOS Regulator Dimethylarginine Dimethylaminohydrolase: A Basis for the Design of Specific Inhibitors

Daniel Frey,[1] Oliver Braun,[1] Christophe Briand,[1] Milan Vašák,[1] and Markus G. Grütter[1,*]
[1] Department of Biochemistry
University of Zürich
Winterthurerstrasse 190
CH-8057 Zürich
Switzerland

Summary

Dimethylarginine dimethylaminohydrolase (DDAH) is involved in the regulation of nitric oxide synthase (NOS) by metabolizing the free endogenous arginine derivatives N^{ω}-methyl-L-arginine (MMA) and N^{ω},N^{ω}-dimethyl-L-arginine (ADMA), which are competitive inhibitors of NOS. Here, we present high-resolution crystal structures of DDAH isoform 1 (DDAH-1) isolated from bovine brain in complex with different inhibitors, including *S*-nitroso-L-homocysteine and Zn^{2+}, a regulator of this mammalian enzyme. The structure of DDAH-1 consists of a propeller-like fold similar to other arginine-modifying enzymes and a flexible loop, which adopts different conformations and acts as a lid at the entrance of the active site. The orientation and interaction mode of inhibitors in the active site give insight into the regulation and the molecular mechanism of the enzyme. The presented structures provide a basis for the structure-based development of specific DDAH-1 inhibitors that might be useful in the therapeutic treatment of NOS dysfunction-related diseases.

Introduction

Nitric oxide (NO) is an important signaling and effector molecule involved in pathophysiological and physiological processes such as neurotransmission, regulation of vascular tone, and bacterial defense (Colasanti and Suzuki, 2000). NO is produced by three homologous nitric oxide synthases (NOS) converting L-arginine to L-citrulline: endothelial NOS (eNOS) and neuronal NOS (nNOS), which are predominantly regulated by cytosolic Ca^{2+}, which, in turn, modulates protein-protein interaction between calmodulin and NOS (Bredt and Snyder, 1994; Kone et al., 2003), and the inducible NOS (iNOS). High toxicity, free diffusibility, and potent chemical reactivity of the free radical NO require strict, tight control of the three isoforms of NOS to enable efficient signaling and to prevent simultaneous damage to other cellular components (Rassaf et al., 2004). For example, several cardiovascular risk factors like hypertension, smoking, diabetes mellitus, and vascular inflammation are related to the dysfunction of NOS (Vallance and Leiper, 2004).

It is well established that NO synthesis can be competitively inhibited by the endogenously methylated arginine derivatives N^{ω}-methyl-L-arginine (monomethylated L-arginine, MMA) and N^{ω},N^{ω}-dimethyl-L-arginine (asymmetric dimethylated L-arginine, ADMA). MMA and ADMA are competitive endogenous inhibitors of all three isoforms of NOS, while N^{ω},$N^{\omega'}$-dimethyl-L-arginine (symmetric dimethylated L-arginine, SDMA) does not inhibit NOS (Leiper and Vallance, 1999; Tran et al., 2003). Methylated arginines appear in proteins that are modified posttranslationally (Paik and Kim, 1971) or as free amino acids in various mammalian tissues (Ueno et al., 1992). The enzyme N^{ω},N^{ω}-dimethyl-L-arginine dimethylaminohydrolase (DDAH, EC 3.5.3.18) converts MMA and ADMA to L-citrulline and mono- or dimethylamine, respectively, thus regulating NOS activity (MacAllister et al., 1996). This enzyme does not degrade the non-NOS inhibitor SDMA.

Mammalian DDAH exists in two isoforms—DDAH-1, which colocalizes mainly with nNOS, and DDAH-2, which is predominantly expressed in tissues containing eNOS—suggesting an isoform-dependent regulation of NOS (Tran et al., 2000). Overexpression of DDAH activates NOS (Dayoub et al., 2003), whereas inhibition of DDAH leads to accumulation of ADMA, causing decreased NOS activity and NO levels (Cooke, 2004). Increased levels of MMA and ADMA in plasma and urine have been implicated in diseases linked with low levels of NO, such as uremia, chronic heart failure, atherosclerosis, and hyperhomocysteinemia (Tsao and Cooke, 1998). In contrast, in diseases such as septic shock, migraine, inflammation, and neurodegenerative disorders, excess NOS activity or NO production is involved (Vallance and Leiper, 2002). Attenuation of NO production by inhibiting DDAH might therefore be a valuable approach to fight these diseases.

Besides the known endogenous DDAH inhibitors L-citrulline, Zn^{2+}, L-homocysteine, and *S*-nitroso-L-homocysteine, only a few designed inhibitors that are either not selective for DDAH or lack DDAH isoform specificity are currently known (Rossiter et al., 2005; Stone et al., 2005; Vallance et al., 2005). Although the crystal structure of a bacterial DDAH is known (Murray-Rust et al., 2001), modeling of a mammalian enzyme structure was deemed difficult because of low sequence similarity (Rossiter et al., 2005).

Here, we present various high-resolution crystal structures of the mammalian DDAH-1 from bovine brain, which include: (1) structures of the unliganded enzyme in two different crystal forms, (2) structures of the Zn^{2+} bound enzyme at two different pH values, and (3) structures of the enzyme in complex with the reaction product L-citrulline, the inhibitor L-homocysteine, and the endogenous suicide inhibitor *S*-nitroso-L-homocysteine. Based on the DDAH-1 structure, a high-quality homology model of the bovine enzyme isoform DDAH-2 could be built that reveals differences in the binding pocket between both isoforms.

The structures presented here can serve as an ideal basis for the design of inhibitors that could have a potential pharmaceutical use in diseases in which overproduction of NO is involved (Vallance and Leiper, 2002).

*Correspondence: gruetter@bioc.unizh.ch

Table 1. X-Ray Data and Refinement Statistics

	Zn^{2+} (pH 9.0)	Zn^{2+} (pH 6.3)	I[a]	II[a]	Cit[b]	Hcy[c]	HcyNO[d]
Space group	$P2_12_12_1$	$P2_12_12_1$	$P2_1$	$P2_1$	$P2_1$	$P2_1$	$P2_1$
Wavelength (Å)	1.0000	0.8550	0.9793	0.9793	1.0001	0.9793	0.8998
Cell dimensions							
a (Å)	44.38	44.40	40.96	42.84	40.7	44.7	40.66
b (Å)	74.63	76.49	79.61	81.02	80.1	79.4	79.91
c (Å)	81.40	81.49	44.87	44.68	44.9	70.6	44.89
β (°)	90	90	108.5	100.6	108.08	104.8	108.58
Resolution (Å)	1.6	2.0	1.7	1.7	1.2	1.8	1.08
Completeness (%)[e]	98.4 (96.0)	99.2 (96.3)	96.2 (93.3)	99.3 (93.6)	98.0 (94.3)	98.1 (90.7)	98.6 (96.0)
Number of unique reflections	35,849	19,299	28,708	32,693	91,611	48,415	114,012
Multiplicity	7.24	5.08	2.62	3.45	2.98	3.25	3.65
$I/\sigma(I)$[e]	10.4 (4.5)	10.8 (4.0)	12.0 (4.8)	16.9 (4.8)	10.8 (2.0)	9.5 (2.2)	15.7 (4.1)
R_{sym}[e]	7.6 (49.7)	6.8 (44.2)	8.7 (29.3)	6.0 (19.8)	7.0 (33.4)	12.6 (37.6)	7.7 (33.9)
Processing program	XDS	Denzo	Denzo	Denzo	Denzo	Denzo	Denzo
Resolution range (Å)	20–1.6	40–2.0	30–1.7	30–1.7	20–1.2	30–1.8	20–1.08
Number of reflections in working and test set	35,844/1,182	19,241/935	27,225/1,392	31,033/1,622	81,946/2,580	45,805/2,422	112,848/1,126
R_{cryst} (%)	20.3	22.9	18.8	19.4	12.2	17.6	11.1
R_{free} (%)	23.8	25.9	21.9	21.7	16.8	21.2	14.0
Water molecules[f]	333	98	336	247	393 (5)	656	482
Rmsd							
Bond lengths (Å)	0.004	0.006	0.004	0.004	0.013[g]	0.007	0.016[g]
Bond angles (°)	1.3	1.2	1.4	1.3	0.030[g]	1.4	0.034[g]
Mean B value ($Å^2$)	24.0	29.6	16.6	21.0	15.0	14.0	13.3

[a] Crystal forms I and II of Zn^{2+} free DDAH-1.
[b–d] Complex structures with L-citrulline (Cit), L-homocysteine (Hcy), and *S*-nitroso-L-homocysteine (HcyNO), respectively.
[e] Values in parentheses are for the highest-resolution shell.
[f] Numbers in parentheses indicate the number of half waters.
[g] Rmsd values in Å for the refinements with SHELX.

Results

Overall Structure

We determined crystal structures of ligand-free DDAH-1 in two different crystal forms, of DDAH-1 with bound Zn^{2+} at pH 6.3 and pH 9.0, and of DDAH-1 in complex with L-citrulline, L-homocysteine, and *S*-nitroso-L-homocysteine at high resolution (between 1.08–2.0 Å; see Table 1). The overall fold of the enzyme consists of five repeats of a ββαβ motif (Figure 1A) with a root-mean-square deviation (rmsd) of 0.19–0.50 Å for all Cα atoms between the different structures. The CATH Database classifies the protein as a five-stranded propeller belonging to the homologous superfamily of L-arginine/glycine amidinotransferase (3.75.10.10) (Orengo et al., 1997).

As a result of the modular architecture of DDAH-1, there is a channel in the center of the molecule that is closed in the middle by a salt bridge between Glu77 and Lys174 that, in turn, forms the bottom of the active site (Figure 1B). One side of the channel represents a water-filled pore delineated by the first β strand of each of the five propeller blades. The other side of the channel is the active site cleft whose outermost boundaries are defined by short loop regions and α helical structures. A flexible loop region formed by amino acids 25–36 closes the active site upon binding of the substrate (Figures 1A and 2).

Structure of the Unliganded Enzyme and Conformational Changes in the Lid Region

Three different conformations for a surface stretch representing amino acids 25–36, here referred to as the lid region, were observed in the various structures (Figure 2). A closed form was observed in five structures of the liganded enzyme, in the Zn^{2+}-containing structure at pH 9.0, and in one unliganded enzyme structure (crystal form I, see Table 1). Crystal packing in crystal form I containing the unliganded enzyme prevents the lid region from adopting an open conformation. The open conformation was observed in crystal form II of the unliganded enzyme. Here, the lid region forms an α helix reorienting Leu29 12 Å away from the active site and leaving the active site entrance open and accessible for the substrate to enter and bind. An alternative open conformation was observed in the Zn^{2+}-containing structure at pH 6.3, indicating high flexibility of the lid region in the open conformation. The high flexibility of the lid region is also reflected by high crystallographic B factors: the average B factor value for amino acids 25–36 is 50 $Å^2$, compared to an overall value of 21 $Å^2$ (Figure 3). The closed lid conformation allows for the formation of an additional hydrogen bond between the amino group of the ligand and the main chain carbonyl of Leu29. This additional interaction enables the rearrangement from the α-helical (open) to the loop (closed) conformation of the lid. In the latter case, the isobutyl side chain of Leu29 blocks the entrance of the active site (Figure 2).

Structures of Zn^{2+} Bound to the Active Site

The protein purified from bovine brain contained one Zn^{2+} ion with an occupancy of ∼95% and an apparent dissociation constant of ∼4 nM as previously described (Knipp et al., 2001). This protein was crystallized at pH 6.3 and pH 9.0. The conformation of the lid loop is different in the two structures. In the crystal form grown at pH

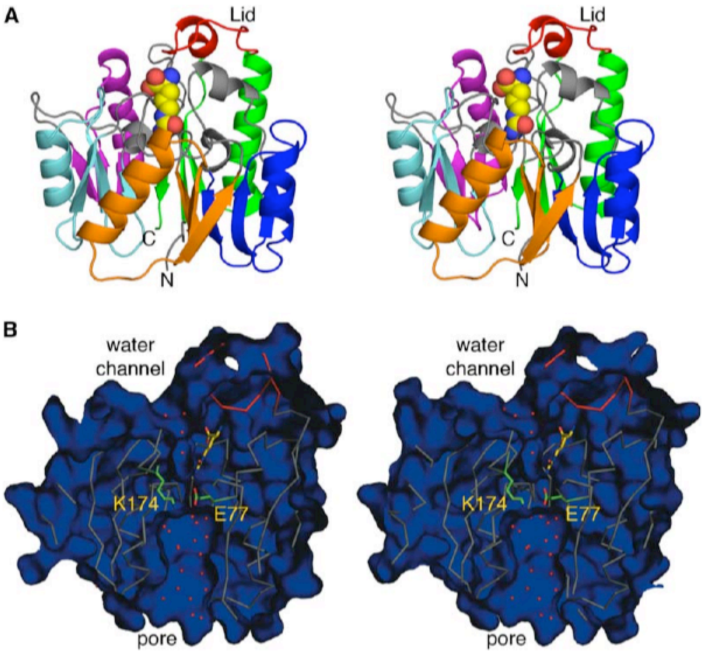

Figure 1. Overall Structure of DDAH-1

(A) Stereoview of the overall structure of DDAH-1 (ribbon depiction) in complex with L-citrulline (space filled). The ββαβ blades of the propeller are colored green, magenta, cyan, orange, and blue from the N to the C terminus. The connecting loops and helices between the blades are colored gray. Residues 25–36 form a lid shown in red. L-citrulline is shown with carbon atoms in yellow, nitrogen in blue, and oxygen in red. All figures were generated with the program PyMOL (DeLano, 2002).

(B) Stereoview of a cross-section through DDAH-1. The DDAH-1 L-citrulline complex was used for the surface representation. Water molecules were excluded for the calculation of the surface. Water molecules in the water channel and in the pore are shown as red spheres, and L-citrulline is shown as a stick model. The two residues, Glu77 and Lys174, shielding the active site from the water-filled pore are shown in green, and the lid is shown in red.

6.3, the lid region adopts a second open conformation (Figure 2) that is fixed by water molecules mediating crystal contacts. At pH 9.0, the lid is in a closed conformation that is induced by a glycine molecule present in the crystallization buffer acting as a ligand.

The observed distances and angles between zinc and the coordinating atoms at both pH values are summarized in Tables 2 and 3. In both structures, the Zn^{2+} is bound to the enzyme in a distorted tetrahedral environment coordinated by Cys273 and three water molecules. At pH 6.3, two water molecules are hydrogen bonded to the side chains of Asp78 and Glu77. The third water molecule is not stabilized by additional protein ligands (Figure 4A and Tables 2 and 3).

At pH 9.0, the Zn^{2+} is still in a tetrahedral environment; however, the position compared to the structure at pH 6.3 is slightly changed (Figure 4B and Tables 2 and 3). At pH 9.0, the side chain of His172 is deprotonated and the salt bridge to Asp126 is weakened, allowing His172 to adopt two different conformations in the same structure (Figure 4B). The first conformation is equivalent to

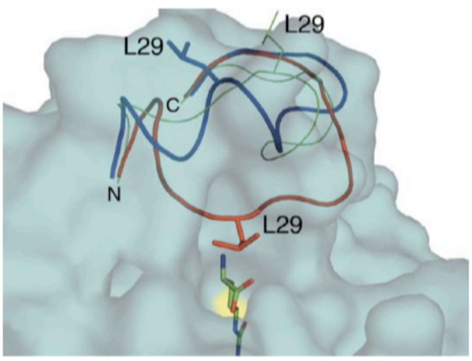

Figure 2. Conformations of the Lid Region

The three different conformations of the lid region of DDAH-1 are shown. The open conformation (crystal form II), the closed conformation (crystal form I), and the Zn^{2+} bound structure at pH 6.3 are shown in blue, red, and green, respectively (Table 1). Residues 24–37 for all three structures are represented as ribbons, and the side chain Leu29 is shown as a stick model. The surface is calculated by using the structure of crystal form II without water molecules. L-citrulline is shown as a stick model, and the active site cysteine is marked by a yellow surface. The distance between the Cγ atom of Leu29 in the open and closed conformation is 12 Å. N and C mark the N- and C-terminal end of the lid, respectively.

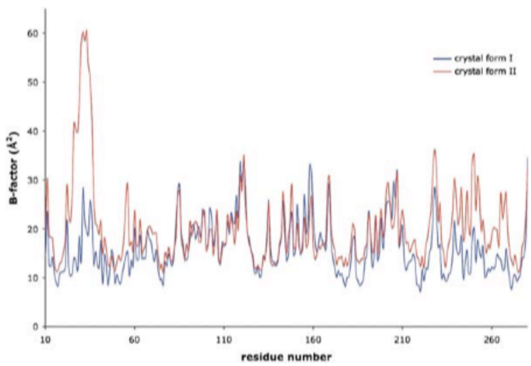

Figure 3. Comparison of the Temperature Factors of Crystal Forms I and II

B factors ($Å^2$) of each residue were averaged and plotted against the residue number. Overall B factors of crystal forms I and II were 17 and 21 $Å^2$, respectively. A mean B value of 50 $Å^2$ in the open conformation (crystal form II) indicates the high mobility of the lid region. In the closed conformation (crystal form I), the mean B value is only 18 $Å^2$ for these residues, indicating that the lid region is more rigid.

Table 2. Distances between Zinc and the Coordinating Atoms

Crystal Structure	Zn-S(p)[a]	Zn-N(p)[a]	Zn-O(w)[a]
pH 6.3[b]	2.19	—	1.84, 1.92, 2.37
pH 9.0[c]	2.27	2.38	2.08, 2.03, 1.81
Mean value of observed distances (Alberts et al., 1998)	2.21 ± 0.13	2.07 ± 0.09	2.12 ± 0.15

[a] (p) or (w) indicate protein or water ligands, respectively.
[b] Coordinate error ± 0.23.
[c] Coordinate error ± 0.1 Å (Cruickshank, 1999).

that seen at pH 6.3. Cys273 and three water ligands, which are hydrogen bonded to Asp78, Glu77, and His172, coordinate the Zn^{2+} ion. In the second conformation, the imidazole group coordinates the Zn^{2+} directly: the distal nitrogen of the side chain is now in the position as the water ligand observed in the first conformation.

L-Citrulline, L-Homocysteine, and *S*-Nitroso-L-Homocysteine Bound to DDAH-1

Crystals of the closed lid conformation were obtained with L-citrulline, the product of the enzymatic reaction bound in the active site of DDAH-1 (Figure 5A). Difference electron density maps revealed clear density in the active site for the ligand, although citrulline was not included in the first round of model building (Figure 5A). Its Cα-carboxyl group forms three salt bridges, two with the guanidino group of Arg144 of DDAH-1, and one with the guanidino group of Arg97. Two hydrogen bonds between the Cα-amino group of the ligand and the main chain carbonyls of Val267 and Leu29 of the enzyme as well as a salt bridge to the side chain of Asp72 further stabilize L-citrulline in the active site. In addition, side chains of Asp78 and Glu77 form hydrogen bonds to the ureido moiety of L-citrulline and position the ligand in the active site. The hydrophobic moiety (Cβ–Cδ) is shielded by Phe75 on one side and Leu171 on the other side. Sγ of Cys273 points away from L-citrulline, indicating a steric hindrance between the protein and the ligand. The active site water occupies the same position as it did in the structure of the bacterial enzyme (Murray-Rust et al., 2001). It forms a hydrogen bond with the side chain hydroxyl group of Ser175 and the proximal nitrogen of the imidazole ring of His172. The active site water molecule is linked to a series of three water molecules in a water channel leading to the surface of the enzyme (Figure 1B). This water channel allows for the replenishment of water molecules that participate in the hydrolysis of MMA and ADMA.

L-homocysteine has been identified recently as a risk factor for cardiovascular diseases (Lentz et al., 2003). A number of investigations report on elevated levels of ADMA and a decreased production of NO in hyperhomocysteinemia (Böger et al., 2000; Selley, 2003; Stühlinger et al., 2003). Recently, homocysteine was reported to directly inhibit DDAH in vitro. From their data, the authors proposed the formation of a disulfide bridge between homocysteine and DDAH (Stühlinger et al., 2001).

Our structure in the presence of L-homocysteine clearly indicates that L-homocysteine binds in the active site pocket in the same orientation as L-citrulline (Figure 5B), and that it binds with a closed lid conformation. However, even in the absence of reducing agents, no electron density of a formed disulfide bond was found between the two sulfur atoms. Therefore, our results do not support the formation of a disulfide bond proposed by Stühlinger et al. (2001).

Knipp et al. (2005) showed that upon reaction of DDAH-1 with *S*-nitroso-L-homocysteine a so far unobserved covalent product, *N*-thiosulfoximide (Cys273:S-NH-S(O):homocysteine), exists in the active site. The proposed reaction proceeds through a sulfiliminosulfonium ion (Cys273:S-N=S^+-homocysteine ↔ Cys273:S^+=N-S-homocysteine). Isotope labeling experiments revealed that the oxygen atom from a water molecule is incorporated, reacting exclusively with the sulfur atom of homocysteine (Knipp et al., 2005). In the crystal structure, *N*-thiosulfoximide is clearly seen in the difference Fourier electron density map (Figure 5C). Interestingly, refinement of the structure revealed two isomers of *N*-thiosulfoximide: Cys273:S-NH-S(O):homocysteine and Cys273:S(O)-NH-S:homocysteine. Since the specific attack of a water molecule on the sulfur atom of homocysteine has already been shown (Knipp et al., 2005), the formation of both isomers in the crystal structure is, to our knowledge, new, indicating a subsequent slow rearrangement previously not observed.

N-thiosulfoximide is stabilized by several salt bridges and hydrogen bonds (Figure 6). Similar to what occurs in the other DDAH-1 complex structures, Arg144 and Arg97 stabilize the Cα-carboxyl group via salt bridges, whereas Leu29, Val267, and Asp72 stabilize the Cα-amino group. As was already observed in the DDAH-1 L-citrulline complex, the lid region adopts a closed conformation. The nitrogen atom of *N*-thiosulfoximide forms a hydrogen bond to Asp78 in both isomers. However, the hydrogen bond between the thionyl oxygen and the proximal nitrogen atom of the imidazole group of His172 is only present in the isomer with the oxygen atom bound to the homocysteine sulfur atom. On the other hand, a hydrogen bond to the main chain nitrogen of Leu270 stabilizes the isomer with the oxygen atom located at the Cys273 sulfur atom.

The nitrogen atom of *N*-thiosulfoximide occupies a similar position as Nε of L-citrulline, and there is a slight shift of 0.6 Å toward Cys273 due to longer S-N and C-S bonds and the missing repulsion of the free side chain of Cys273 as seen in the DDAH-1 L-citrulline complex. As

Table 3. Angles between Zinc and the Coordinating Ligands

Crystal Structure	S-Zn-N	N-Zn-O	S-Zn-O	O-Zn-O
pH 6.3	—	—	112.0, 107.6, 100.1	102.5, 109.3, 125.0
pH 9.0	126.7	99.6, 105.6	113.9, 114.8, 110.0	114.0, 95.6, 107.3
Mean value of observed angles (Alberts et al., 1998)	112 ± 4	110 ± 9	105 ± 7	111 ± 9

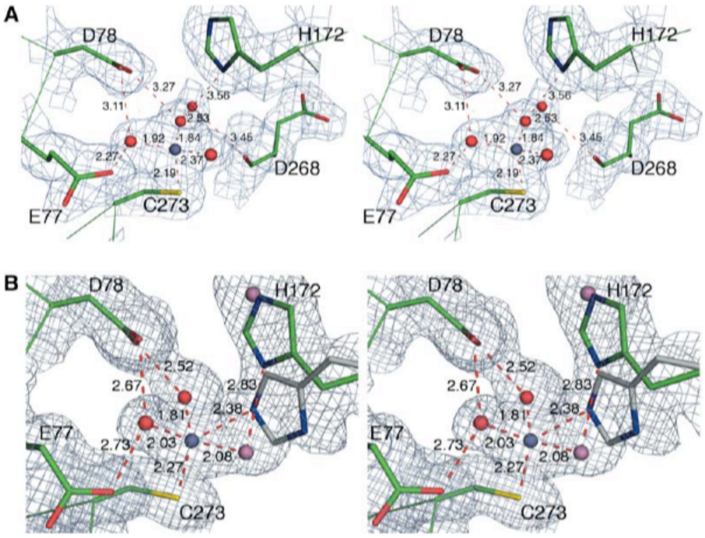

Figure 4. Zn^{2+} Binding Site

(A) $2F_o - F_c$ Fourier map of the Zn^{2+} binding site at pH 6.3. The active site residues Cys273, His172, Glu77, and Asp78 and residue Asp268 are shown as sticks. The bound Zn^{2+} ion is shown in gray, and the bound water molecules are shown in red. Metal ion ligand interactions are represented as dotted lines, and their corresponding lengths are indicated in Å.

(B) $2F_o - F_c$ Fourier map of the Zn^{2+} binding site at pH 9.0. Interactions of the metal ion in the first and second coordination sphere are shown. The Zn^{2+} atom is shown in gray, and water molecules are shown in red. The second conformation of His172 and the alternating water molecule are shown in gray and purple, respectively.

structural information on *N*-thiosulfoximides are not available in the Cambridge Structural Database (Allen, 2002), bond lengths were compared to similar compounds. Bond lengths of 1.81, 1.62, 1.63, 1.5, and 1.81 Å for Cβ to Sγ, Sγ to N, N to Sδ, Sδ to O, and Sδ to Cγ (Cys273:S-NH-S(O):Hcy), respectively, agree well with modeled data for oxidized methionine (Schuttelkopf and van Aalten, 2004) and small-molecule X-ray data for sulfoximines (Cambridge Structural Database [Allen, 2002]). As seen for small molecules, the S-N bond lengths in the *N*-thiosulfoximide isomers are shorter (1.62/1.63 ± 0.03 Å and 1.59/1.62 ± 0.03 Å respectively [Cruickshank, 1999]) than expected for a single bond S-N (1.7 Å) indicating partial double bond character. Remarkably, superposition of the active site residues of the L-citrulline and *N*-thiosulfoximide structures only showed a difference in the newly formed S-N bond of the *N*-thiosulfoximide. As the first collection of diffraction data revealed a partially cleaved *N*-thiosulfoximide resulting in several not clearly defined degradation products in the active site, e.g., cysteine or homocysteine derivatives of sulfenic acids (R-S-OH) or sulfinic acids (R-S(=O)-OH), further data acquisition was performed with lower radiation intensity (data not shown).

Besides the formation of *N*-thiosulfoximide, an *S*-nitrosylation of Cys83 was also observed. The solvent-exposed Cys83 is located at the lower end of the pore. This modification may be due to a transnitrosylation reaction between *S*-nitroso-L-homocysteine and Cys83 under the crystallization conditions. The previously reported *S*-nitrosylation of Cys221 with DEA/NO, liberating gaseous NO, was not observed in our structure (Knipp et al., 2003). Cys221 is located in the core of the protein and is accessible to gaseous NO, whereas modifications by *S*-nitrosohomocysteine are restricted to solvent-accessible residues like Cys273 or Cys83.

Comparison to Other Arginine-Modifying Enzymes

The overall fold of DDAH-1 from bovine brain is identical to the fold of *Pseudomonas aeruginosa* DDAH (*Pa* DDAH; rmsd 1.39 Å) and to other arginine- or substituted arginine-modifying enzymes of which crystal structures have been determined. Peptidylarginine deiminase (PAD4) (Arita et al., 2004), converting L-arginine to L-citrulline and ammonium; the Arg:Gly amidinotransferase (AT) (Humm et al., 1997), involved in creatine biosynthesis; and arginine deiminase (ADI) (Das et al., 2004), an enzyme that uses arginine as an energy source in the arginine-deiminase pathway, showed rmsd values of 1.9, 1.78, and 1.82 Å, respectively.

The sequence identity between DDAH-1 and *Pa* DDAH is low, around 30%. Superposition of the two structures reveals differences in the length and positions of α helices. The active site clefts nevertheless are very similar. Except for Asp126, all amino acids involved in binding and conversion of the substrate ADMA to L-citrulline and dimethylamine are identical in both structures. Remarkably, amino acids 25–36 of DDAH-1 and amino acids 15–25 of *Pa* DDAH showed no sequence identity other than a conserved Leu29 or Leu18 in *Pa* DDAH, demonstrating the importance of this amino acid for the function of the lid (Figure 7).

Superposition of DDAH-1 in complex with L-citrulline and ADI (crystal form I [Das et al., 2004]) revealed a similar positioning of the ureido moiety of L-citrulline near the active site cysteine. Compared to the covalent intermediate structure of ADI, the ideal position for the carbon atom of the substrates to be attacked resides 0.4 ± 0.04 Å (Cruickshank, 1999) closer to the active site cysteine. Although the side chain of Cys273 rotates 120° around Cβ to overcome the close contact of Sγ with L-citrulline, we ascribe the small shift of the L-citrulline molecule to these repulsive interactions.

Lys174 of DDAH-1 is a key determinant for substrate specificity. The positively charged side chain of Lys174 in DDAH-1 is unable to coordinate the substrate ADMA or MMA. Instead, Glu77 of DDAH-1 coordinates the unmethylated nitrogen of the substrate. This enables the enzyme to accommodate the bulkier substrates MMA and ADMA. ADI and AT contain an Asp at this position that coordinates both terminal nitrogens in the guanidino moiety of arginine. Furthermore, the pockets that

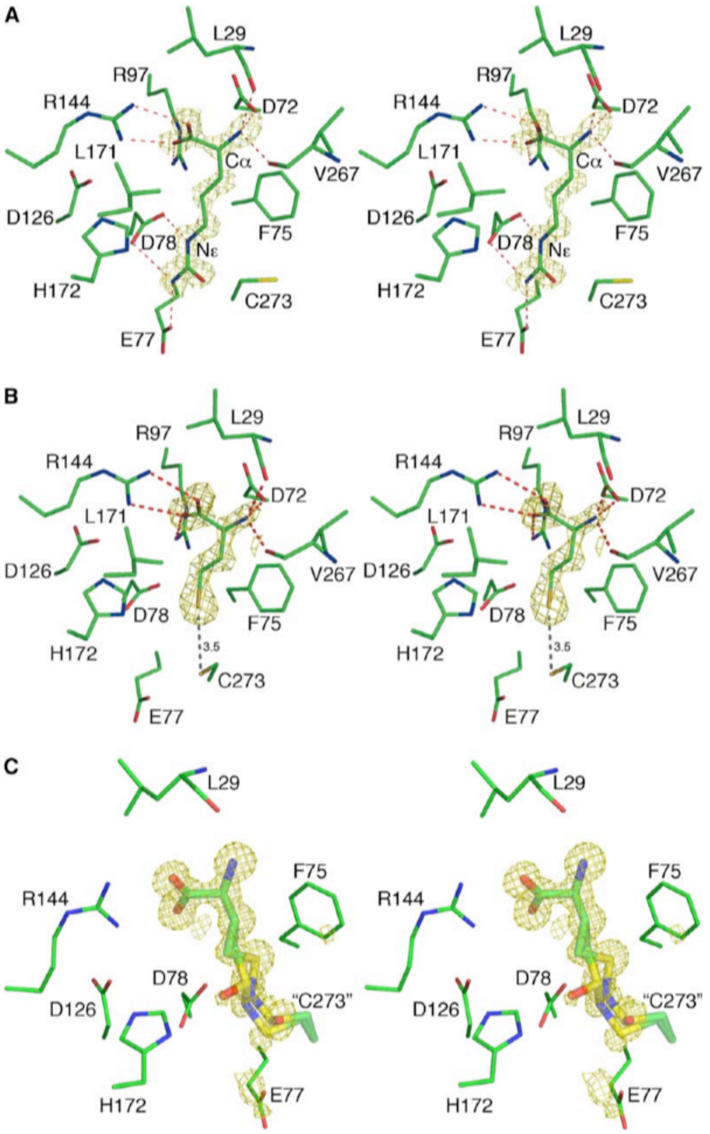

Figure 5. Active Site of DDAH-1 with Bound Ligand

(A–C) Stereoviews of the active site of DDAH-1 with (A) L-citrulline, (B) L-homocysteine, and (C) *S*-nitroso-L-homocysteine are shown. Difference electron density maps shown in yellow and contoured at 3σ were calculated by using the model of DDAH-1 without ligand. The ligand protein interactions are shown as red lines. Sγ of Cys273 and Sδ of homocysteine are 3.5 Å apart (black, [B]).

bind the methylamine or dimethylamine moiety of the substrates in DDAH-1 and *Pa* DDAH are larger and are composed of polar and hydrophobic residues, whereas ADI has a smaller pocket, which is formed only by polar side chains. Despite the differences mentioned, comparison of the structures supports the finding that members of the family of guanidino group-modifying enzymes share a common catalytic mechanism (Shirai et al., 2001).

Comparison of the Two Isoforms of DDAH

Analysis of human DDAH mRNA distribution in various tissues indicates differential expression of both isoforms in brain regions, in immune cells, and during development. Mainly, DDAH-2 was found in immune cells that express iNOS, suggesting the importance in the control of MMA and ADMA levels (Tran et al., 2000). The alignment of three mammalian DDAH protein sequences is shown in Figure 7. Percentage scoring values of 92% for DDAH-1 and 95% for DDAH-2 protein sequences show high identity within an isoform, whereas the score drops to ∼50% when sequences between the two isoforms are compared. Nevertheless, residues directly involved in the catalytic mechanism (Cys273, Asp126, and His172 in DDAH-1 and Cys276, Asp125, and His171 in DDAH-2) are conserved throughout the compared mammalian DDAH sequences. Variations in amino acid sequences are found in substrate binding residues and in the lid region. The lid region in DDAH-2 is highly conserved and negatively charged through Glu27, whereas proteins belonging to the DDAH-1 isoform contain a less conserved, positively charged lid region. The amino acids that are directly involved in substrate binding are conserved within each isoform among different species, but they are altered between isoforms 1 and 2. These different direct interactions include hydrophobic

Crystal Structure of Mammalian DDAH
907

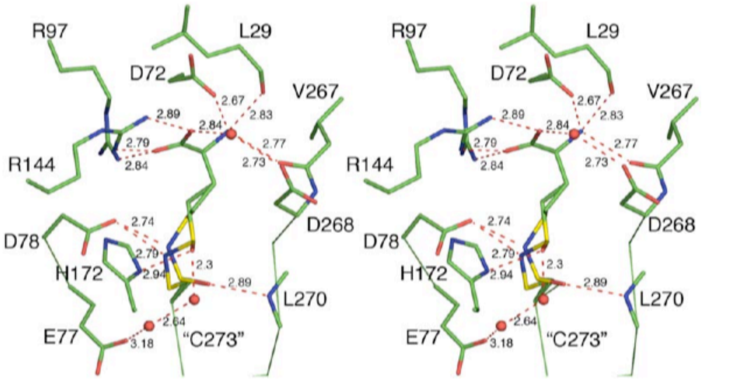

Figure 6. *N*-Thiosulfoximide in the Active Site of DDAH-1

Stereoview of the binding of *N*-thiosulfoximide to DDAH-1. The water molecules involved in binding of the inhibitor are shown as red spheres. The connecting hydrogen bonds and salt bridges are colored as red dots, and their corresponding lengths are indicated in Å. Nitrogen atoms are blue, oxygen atoms are red, and sulfur atoms are yellow.

contacts, e.g., Phe75 to Leu74 in DDAH-2, as well as polar interaction, e.g., Glu77 to Gly76 or Arg144 to Trp143. The modeled active site of DDAH-2 superimposed on the experimentally determined active site of DDAH-1 illustrates these differences (Figure 8). The exchange of Ala28 for Glu27 in DDAH-2 might compensate for the Asp72-to-Leu71 mutation because both acidic side chains are ideally positioned to form salt bridges to the Cα-amino group of the substrate, whereas the distance of the side chain of Trp143 is too far away to coordinate the Cα-carboxyl of the substrate. Overall, three salt bridges between the enzyme and the ligand are altered between the two isoforms. Although there are differences in the binding of the substrate to the enzyme, the overall architecture of DDAH-2 is similar to the one observed in the DDAH-1 structures.

Discussion

Despite the low sequence identity between the mammalian DDAH-1 and the bacterial *Pa* DDAH (Murray-Rust et al., 2001), the overall fold is identical and the active sites are very similar. However, significant differences exist that are important for a better understanding of the enzyme mechanism and critical for structure-based drug design. Our study of the mammalian enzyme

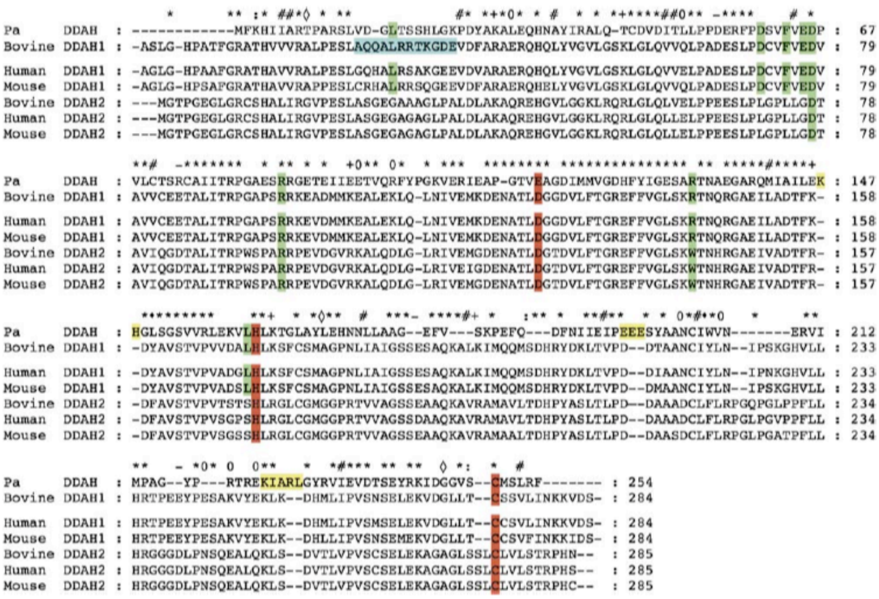

Figure 7. Alignment of DDAH Sequences

The sequence alignment of mammalian DDAH sequences and the bacterial *Pa* DDAH sequence (top, based on bacterial and mammalian enzyme structures). The line above the sequences indicates identical, asterisk; positively charged, plus sign; negatively charged, minus sign; differently charged, "0"; tiny polar, colon; tiny nonpolar, open diamond; aliphatic, pound sign; and aromatic residues, closed diamond in mammalian DDAH sequences. The green bars indicate residues involved in binding of L-citrulline, red residues indicate residues involved in conversion of the substrate, blue indicates amino acids in the lid region, and yellow indicates residues with a large deviation in Cα position between DDAH-1 and *Pa* DDAH. ClustalW (Thompson et al., 1997) was used for the mammalian sequence alignments, whereas *Pa* DDAH was structurally aligned to DDAH-1 (crystal form I).

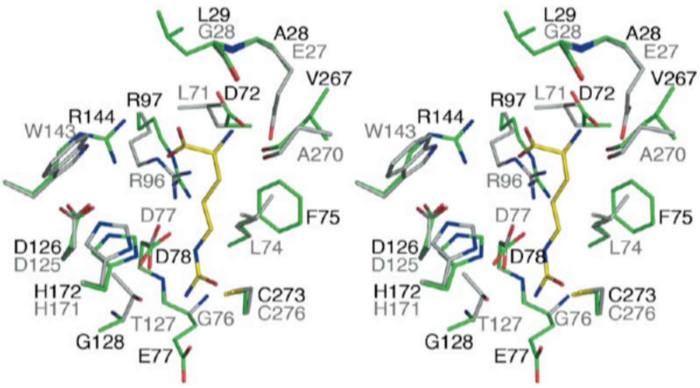

Figure 8. Comparison of DDAH-1 and DDAH-2

Stereoview of the superimposed active sites of the experimentally determined structure of DDAH-1 (green) and the modeled DDAH-2 (gray). L-citrulline is shown in yellow. The main chain, except for glycines and carbonyl atoms involved in binding, is hidden. The model was created with the program Modeller 8v1 (Marti-Renom et al., 2000).

shows that the lid adopts a helical structure and an open conformation when no ligand or only Zn^{2+} is bound, and that it closes the active site when L-citrulline, L-homocysteine, or *S*-nitroso-L-homocysteine are bound. The isobutyl side chain of Leu29 in the lid blocks the entrance to the active site completely, buries the ligands, and thereby stabilizes the enzyme ligand complex. After hydrolysis of MMA or ADMA, both products, L-citrulline and methylamine or dimethylamine, are released from the reaction center. L-citrulline leaves the enzyme upon the opening of the closing lid through the active site, whereas, for both amines, two possibilities exist. First they could leave in a manner similar to L-citrulline—through the active site entrance. Alternatively, they could leave through the pore opened by small movements of the side chains of Glu77 and Lys174.

As isolated, the enzyme contains one tightly bound zinc ion (K_d of $\sim$4 nM) that reversibly inhibits its enzymatic activity. Knipp et al. (2001) reported that the coordination sphere of this Zn^{2+} ion contains two sulfur atoms and two lighter atoms (N or O). Coordination of Zn^{2+} by Sγ of Cys221 and Cys273 was proposed since Zn^{2+} protected both sulfurs from *S*-nitrosylation by gaseous NO (Knipp et al., 2003). The presented crystal structures of the Zn^{2+} ion containing DDAH-1 demonstrate that besides Cys273 no additional sulfur ligand is found in the proximity of the metal ion. All presented structures adopt the same overall fold, and, besides the loop region, no flexibility of the protein was found; therefore, it can be ruled out that Cys221, which is located 12 Å away, participates in the binding of the metal ion. The second sulfur ligand, observed by EXAFS, must therefore be an exogenous ligand, introduced during purification of the enzyme. The other ligands of the Zn^{2+} ion observed in the structures are water molecules and His172. Since the second conformation of His172 is only present in the structure at pH 9, we suggest that this conformation rather reflects a high pKa value of this histidine that is similar to the pKa values of histidines in the active sites of other cysteine proteases (Lakhdar-Ghazal et al., 2002). Water molecules positioned by the protein as in the active site at pH 6.3 or pH 9.0 mediate tight binding equal to that of a direct protein ligand; therefore, we assume that the metal ion is coordinated by one protein ligand (Cys273) and three water molecules, which are fixed to the protein by a hydrogen bonding network.

Reduction of NO production is desirable in diseases such as septic shock, pain, arthritis, or asthma (Vallance and Leiper, 2002). In these cases, a specific inhibition of NOS isoforms is sought. So far, the clinical use of NOS inhibitors has been limited due to the absence of specificity or toxicity. Although new promising compounds inhibiting iNOS have already been developed, a complete inhibition of NOS isoforms may be unwanted (Alderton et al., 2005). Inhibition of DDAH isoforms offers an alternative approach to reduce NO levels in a tissue-specific manner. At present, three different classes of DDAH inhibitors, including pentalfluorophenyl sulfonates (Vallance et al., 2005), 2-chloroacetamidine (Stone et al., 2005), and the class of improved DDAH inhibitors, derived from the weak reversible inhibitor *S*-2-amino-4(3-methylguanidino)butanoic acid (4124W) (Rossiter et al., 2005), are reported in the literature. Although, not selective or isoform specific, in combination with the structures presented here, they can serve as a basis for the development of selective/specific inhibitors.

Inhibitors having a distinct selectivity for DDAH should (1) form hydrogen bonds to the side chains of Asp72, Glu77, Asp78, Arg144, and Arg97 and to the main chain carbonyls of Leu29 and Val267, (2) recognize the positively charged side chain of Lys174, and (3) form hydrogen bonds to His172 and the main chain carbonyl of Leu270. These features should provide sufficient affinity to bind to DDAH and selectivity for DDAH compared to binding to ADI or AT. When comparing the active sites of the structure of DDAH-1 with the modeled active site of DDAH-2, seven side chains forming the binding site are different. An inhibitor, which interacts with the indole ring of Trp143 (Arg144 in DDAH-1), should show preferential binding to DDAH-2. A hydrogen bond to Thr127 in DDAH-2 (Gly128 in DDAH-1) would further enhance the specificity toward DDAH-2. In contrast, inhibitors that form a hydrogen bond to the side chain of Glu77 in DDAH-1 (Gly76 in DDAH-2) should bind preferentially to DDAH-1. In addition, the different polarity and size of the side chains of Leu71 (DDAH-2) and Asp72 (DDAH-1) could be exploited to increase the specificity toward one of the isoforms. These differences could allow for the design of selective and specific inhibitors for each isoenzyme. Such inhibitors would be of high interest for DDAH research and treatment of diseases linked to NO overproduction.

Experimental Procedures

Protein Purification

Protein purification and activity measurements were performed as described, with minor modifications (Knipp et al., 2001). All purification steps, except the second ion exchange chromatography, were performed at 4ºC. Tissue homogenization of bovine brains, ammonium sulfate precipitation, and the subsequent hydrophobic interaction chromatography were performed as previously described (Knipp et al., 2001).

Anion Exchange Chromatography

The conductivity of the protein sample was adjusted with H_2O to $\leq$15 mΩ^{-1}cm^{-1} and was then applied to a Fractogel EMD TMAE (s) column (2.6 × 15 cm, Merck) equilibrated with 20 mM triethanolamine/HCl (pH 7.8). The protein was eluted with a linear salt gradient from 0 to 200 mM KCl, and the active fractions were pooled.

Size Exclusion Chromatography

The protein solution was concentrated by ultrafiltration (Amicon) by using a polyethersulfone membrane (Biomax, Milipore) with a molecular weight cutoff of 5 kDa. The final protein sample (10 ml) was applied to a HiLoad 26/60 Superdex 75 column (Pharmacia) equilibrated with 10 mM triethanolamine/HCl, 150 mM KCl (pH 7.4). The active fractions were combined.

Anion Exchange Chromatography

The conductivity of the protein sample was adjusted with H_2O to $\leq$15 mΩ^{-1}cm^{-1} and was then applied to a Mono Q HR 5/5 column (Pharmacia) equilibrated with 10 mM triethanolamine/HCl, 10% (v/v) glycerol (pH 7.4). The protein was eluted with a linear salt gradient from 0 to 200 mM KCl, and the active fractions were pooled.

Protein homogeneity was routinely analyzed by SDS-PAGE and mass spectrometry. Metal-to-protein ratios were determined by measuring the Zn^{2+} concentration by flame atomic absorption spectroscopy (SpectrAA-110, Varian, Inc.), and that of the protein was measured spectrophotometrically (ε_{280} = 14,420 M^{-1} cm^{-1}) (Bogumil et al., 1998).

The fully Zn^{2+}-loaded protein was prepared through protein dialysis against Zn^{2+}-containing buffer (Knipp et al., 2003).

Preparation of Zn^{2+}-free DDAH-1

To remove Zn^{2+} from native DDAH-1, 200 µl protein sample (~50 µM) was dialyzed at 4ºC against 50 ml metal-free 100 mM imidazole/HCl and 10 mM DTT (pH 6.5). The outer solution was steadily bubbled with argon. The dialysis buffer was changed four times (every 2 hr), and dialysis was completed overnight. To remove imidazole and DTT, the protein was dialyzed against two changes of 50 ml metal-free 10 mM MES/NaOH, 150 mM KCl, 2 mM TCEP, 5 mM EDTA (pH 6.5) (Knipp et al., 2001).

Preparation of L-Homocysteine and *S*-Nitroso-L-Homocysteine

L-homocysteine was prepared from L-homocysteine thiolactone·HCl as described by Duerre and Miller (1966) in a nitrogen atmosphere. The yield (96%) was estimated through quantification of thiols (Grasetti and Murray, 1967).

S-nitroso-L-homocysteine was always freshly prepared under anaerobic conditions by using the protocol of Feelisch and Stamler (1996). Briefly, 150 µl 200 mM L-homocysteine was mixed with an equal volume of 200 mM $NaNO_2$. After completion of the reaction, HEPES/EDTA was added to a final concentration of 50 mM HEPES/5 mM EDTA, and the pH was adjusted to 7.3. The concentration of the *S*-nitroso-L-homocysteine formed was determined photometrically with the molar extinction coefficients reported by Feelisch and Stamler (1996).

Crystallization

The protein was concentrated to 2–3 mg/ml in protein storage buffer (10 mM Tris/HCl, 150 mM KCl, 10% [v/v] glycerol [pH 7.4]). Crystallization was done at 4ºC by using the sitting drop vapor diffusion method combining 3 µl protein solution and 3 µl reservoir solution. The final conditions were 100 mM MES/NaOH, 20%–40% PEG 8000 (pH 6.3) or 0.05 M glycine/HCl, 28% PEG 3350 (pH 9.0) for the Zn^{2+}-containing proteins and 100 mM citric acid/NaOH, 20%–40% PEG 8000, 2 mM TCEP (pH 6.3) for the free form of the enzyme and the cocrystallization experiments. For complex structures, 50 mM L-citrulline, 5 mM L-homocysteine, or 0.5 mM *S*-nitroso-L-homocysteine were added to the protein solution prior to crystallization. The crystals were stabilized by adding 20%–25% ethyleneglycol to the crystallization buffer and were flash frozen in liquid propane.

Data Collection, Structure Determination, and Refinement

All data was collected on X06SA at the Swiss Light Source (SLS), Villingen, Switzerland, by using a MarCCD 165 mm detector. Images were processed with XDS (Kabsch, 1993) or Denzo (Otwinoswki and Minor, 1997). Structure determination was solved by molecular replacement by using AMoRe (Navaza, 1994) and a poly-serine model of the DDAH structure from *P. aeruginosa* (PDB code: 1H70). Model building and refinement were done with O (Jones et al., 1991) and CNS (Brunger et al., 1998); the atomic resolution structures (complexes with L-citrulline or *S*-nitroso-L-homocysteine) were further refined with SHELX (Sheldrick and Schneider, 1997). Cocrystallized small molecules were not included in the search models, but clear positive density with contour levels higher than 3σ in difference Fourier maps appeared in the first round of refinement. The position of the Zn^{2+} was confirmed by anomalous difference Fourier maps contoured at 8σ in Fourier maps. The restraints for L-citrulline, the *N*-thiosulfoximide derivative of cysteine, L-homocysteine, and *S*-nitrosocysteine were calculated on the ProDRG server (Schuttelkopf and van Aalten, 2004) and compared with the small-molecule database (Allen, 2002). Data collection and refinement statistics are shown in Table 1. Models were validated by using SFcheck and PROcheck (CCP4, 1994).

Acknowledgments

We gratefully acknowledge the Swiss Light Source, Paul Scherrer Institut, Villigen, Switzerland, for providing synchrotron beamtime and T. Tomizaki, A. Wagner, E. Pohl, and C. Schulze-Briese for their excellent support during data collection. This work was supported by the Swiss National Science Foundation (Grant 6M3100A-1022181 to M.G.G.; Grant 3100A0-100246/1 to M.V.) and the "Hartmann-Müller-Stiftung" (to M.V.).

Received: October 12, 2005
Revised: March 10, 2006
Accepted: March 14, 2006
Published: May 16, 2006

References

Alberts, I.L., Nadassy, K., and Wodak, S.J. (1998). Analysis of zinc binding sites in protein crystal structures. Protein Sci. *7*, 1700–1716.

Alderton, W.K., Angell, A.D., Craig, C., Dawson, J., Garvey, E., Moncada, S., Monkhouse, J., Rees, D., Russell, L.J., Russell, R.J., et al. (2005). GW274150 and GW273629 are potent and highly selective inhibitors of inducible nitric oxide synthase in vitro and in vivo. Br. J. Pharmacol. *145*, 301–312.

Allen, F.H. (2002). The Cambridge Structural Database: a quarter of a million crystal structures and rising. Acta Crystallogr. B *58*, 380–388.

Arita, K., Hashimoto, H., Shimizu, T., Nakashima, K., Yamada, M., and Sato, M. (2004). Structural basis for Ca(2+)-induced activation of human PAD4. Nat. Struct. Mol. Biol. *11*, 777–783.

Böger, R.H., Bode-Böger, S.M., Sydow, K., Heistad, D.D., and Lentz, S.R. (2000). Plasma concentration of asymmetric dimethylarginine, an endogenous inhibitor of nitric oxide synthase, is elevated in monkeys with hyperhomocyst(e)inemia or hypercholesterolemia. Arterioscler. Thromb. Vasc. Biol. *20*, 1557–1564.

Bogumil, R., Knipp, M., Fundel, S.M., and Vašák, M. (1998). Characterization of dimethylargininase from bovine brain: evidence for a zinc binding site. Biochemistry *37*, 4791–4798.

Bredt, D.S., and Snyder, S.H. (1994). Nitric oxide: a physiological messenger molecule. Annu. Rev. Biochem. *63*, 175–195.

Brunger, A.T., Adams, P.D., Clore, G.M., DeLano, W.L., Gros, P., Grosse-Kunstleve, R.W., Jiang, J.S., Kuszewski, J., Nilges, M., Pannu, N.S., et al. (1998). Crystallography & NMR system: a new software suite for macromolecular structure determination. Acta Crystallogr. D Biol. Crystallogr. *54*, 905–921.

CCP4 (Collaborative Computational Project, Number 4) (1994). The CCP4 suite: programs for protein crystallography. Acta Crystallogr. D Biol. Crystallogr. *50*, 760–763.

Colasanti, M., and Suzuki, H. (2000). The dual personality of NO. Trends Pharmacol. Sci. *21*, 249–252.

Cooke, J.P. (2004). Asymmetrical dimethylarginine: the Uber marker? Circulation *109*, 1813–1818.

Cruickshank, D.W. (1999). Remarks about protein structure precision. Acta Crystallogr. D Biol. Crystallogr. *55*, 583–601.

Das, K., Butler, G.H., Kwiatkowski, V., Clark, A.D., Jr., Yadav, P., and Arnold, E. (2004). Crystal structures of arginine deiminase with covalent reaction intermediates; implications for catalytic mechanism. Structure *12*, 657–667.

Dayoub, H., Achan, V., Adimoolam, S., Jacobi, J., Stühlinger, M.C., Wang, B.-y., Tsao, P.S., Kimoto, M., Vallance, P., Patterson, A.J., and Cooke, J.P. (2003). Dimethylarginine dimethylaminohydrolase regulates nitric oxide synthesis: genetic and physiological evidence. Circulation *108*, 3042–3047.

DeLano, W.L. (2002). The PyMOL Molecular Graphics System (San Carlos, CA: DeLano Scientific).

Duerre, J.A., and Miller, C.H. (1966). Preparation of L-homocysteine from L-homocysteine thiolactone. Anal. Biochem. *17*, 310–315.

Feelisch, M., and Stamler, J.S. (1996). Donors of nitrogen oxides. In Methods in Nitric Oxide Research, M. Feelisch and J.S. Stamler, eds. (Chichester, UK: John Wiley & Sons, Ltd.), pp. 71–115.

Grasetti, D.R., and Murray, J.F. (1967). Determination of sulfhydryl groups with 2,2′- or 4,4′-dithiodipyridine. Archives Biochem. Biophys. *119*, 41–49.

Humm, A., Fritsche, E., Steinbacher, S., and Huber, R. (1997). Crystal structure and mechanism of human L-arginine:glycine amidinotransferase: a mitochondrial enzyme involved in cretine biosynthesis. EMBO J. *16*, 3373–3385.

Jones, T.A., Zou, J.Y., Cowan, S.W., and Kjeldgaard. (1991). Improved methods for building protein models in electron density maps and the location of errors in these models. Acta Crystallogr. A *47*, 110–119.

Kabsch, W. (1993). Automatic processing of rotation diffraction data from crystals of initially unknown symmetry and cell constants. J. Appl. Crystallogr. *26*, 795–800.

Knipp, M., Charnock, J.M., Garner, C.D., and Vašák, M. (2001). Structural and functional characterization of the Zn(II)-site in dimethylargininase-1 (DDAH-1) from bovine brain. Zn(II) release activates DDAH-1. J. Biol. Chem. *276*, 40449–40456.

Knipp, M., Braun, O., Gehrig, P.M., Sack, R., and Vašák, M. (2003). Zn(II)-free dimethylargininase-1 (DDAH-1) is inhibited upon specific Cys-*S*-nitrosylation. J. Biol. Chem. *278*, 3410–3416.

Knipp, M., Braun, O., and Vašák, M. (2005). Searching for DDAH inhibitors: *S*-nitroso-L-homocysteine is a chemical lead. J. Am. Chem. Soc. *127*, 2372–2373.

Kone, B.C., Kuncewicz, T., Zhang, W., and Yu, Z.Y. (2003). Protein interactions with nitric oxide synthases: controlling the right time, the right place, and the right amount of nitric oxide. Am. J. Physiol. Renal Physiol. *285*, F178–F190.

Lakhdar-Ghazal, F., Blonski, C., Willson, M., Michels, P., and Perie, J. (2002). Glycolysis and proteases as targets for the design of new anti-trypanosome drugs. Curr. Top. Med. Chem. *2*, 439–456.

Leiper, J., and Vallance, P. (1999). Biological significance of endogenous methylarginines that inhibit nitric oxide synthases. Cardiovasc. Res. *43*, 542–548.

Lentz, S.R., Rodionov, R.N., and Dayal, S. (2003). Hyperhomocysteinemia, endothelial dysfunction, and cardiovascular risk: the potential role of ADMA. Atheroscler. Suppl. *4*, 61–65.

MacAllister, R.J., Parry, H., Kimoto, M., Ogawa, T., Russell, R.J., Hodson, H., Whitley, G.S.J., and Vallance, P. (1996). Regulation of nitric oxide synthesis by dimethylarginine dimethylaminohydrolase. Br. J. Pharmacol. *119*, 1533–1540.

Marti-Renom, M.A., Stuart, A.C., Fiser, A., Sanchez, R., Melo, F., and Sali, A. (2000). Comparative protein structure modeling of genes and genomes. Annu. Rev. Biophys. Biomol. Struct. *29*, 291–325.

Murray-Rust, J., Leiper, J., McAllister, M., Phelan, J., Tilley, S., Santa Maria, J., Vallance, P., and McDonald, N. (2001). Structural insights into the hydrolysis of cellular nitric oxide synthase inhibitors by dimethylarginine dimethylaminohydrolase. Nat. Struct. Biol. *8*, 679–683.

Navaza, J. (1994). AMoRe: an automated package for molecular replacement. Acta Crystallogr. A *50*, 157–163.

Orengo, C.A., Michie, A.D., Jones, S., Jones, D.T., Swindells, M.B., and Thornton, J.M. (1997). CATH—a hierarchic classification of protein domain structures. Structure *5*, 1093–1108.

Otwinoswki, Z., and Minor, W. (1997). Processing of X-ray diffraction data collected in oscillation mode. Methods Enzymol. *276*, 307–326.

Paik, W.K., and Kim, S. (1971). Protein methylation. Science *174*, 114–119.

Rassaf, T., Feelisch, M., and Kelm, M. (2004). Circulating NO pool: assessment of nitrite and nitroso species in blood and tissues. Free Radic. Biol. Med. *36*, 413–422.

Rossiter, S., Smith, C.L., Malaki, M., Nandi, M., Gill, H., Leiper, J.M., Vallance, P., and Selwood, D.L. (2005). Selective substrate-based inhibitors of mammalian dimethylarginine dimethylaminohydrolase. J. Med. Chem. *48*, 4670–4678.

Schuttelkopf, A.W., and van Aalten, D.M. (2004). PRODRG: a tool for high-throughput crystallography of protein-ligand complexes. Acta Crystallogr. D Biol. Crystallogr. *60*, 1355–1363.

Selley, M.L. (2003). Increased concentrations of homocysteine and asymmetric dimethylarginine and decreased concentrations of nitric oxide in the plasma of patients with Alzheimer's disease. Neurobiol. Aging *24*, 903–907.

Sheldrick, G.M., and Schneider, T.R. (1997). SHELXL: high resolution refinement. Methods Enzymol. *277*, 319–343.

Shirai, H., Blundell, T.L., and Mizuguchi, K. (2001). A novel superfamily of enzymes that catalyze the modification of guanidino groups. Trends Biochem. Sci. *26*, 465–468.

Stone, E.M., Person, M.D., Costello, N.J., and Fast, W. (2005). Characterization of a transient covalent adduct formed during dimethylaginine dimethylaminohydrolase catalysis. Biochemistry *44*, 7069–7078.

Stühlinger, M.C., Tsao, P.S., Her, J.-H., Kimoto, M., Balint, R.F., and Cooke, J.P. (2001). Homocysteine impairs the nitric oxide synthase pathway: role of asymmetric dimethylarginine. Circulation *104*, 2569–2575.

Stühlinger, M.C., Oka, R.K., Graf, E.E., Schmolzer, I., Upson, B.M., Kapoor, O., Szuba, A., Malinow, M.R., Wascher, T.C., Pachinger, O., and Cooke, J.P. (2003). Endothelial dysfunction induced by hyperhomocyst(e)inemia: role of asymmetric dimethylarginine. Circulation *108*, 933–938.

Thompson, J.D., Gibson, T.J., Plewniak, F., Jeanmougin, F., and Higgins, D.G. (1997). The CLUSTAL:X windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. Nucleic Acids Res. *25*, 4876–4882.

Tran, C.T.L., Fox, M.F., Vallance, P., and Leiper, J.M. (2000). Chromosomal localization, gene structure, and expression pattern of *DDAH1*: comparison with *DDAH2* and implications for evolutionary origins. Genomics *68*, 101–105.

Tran, C.T.L., Leiper, J.M., and Vallance, P. (2003). The DDAH/ADMA/NOS pathway. Atheroscler. Suppl. *4*, 33–40.

Tsao, P.S., and Cooke, J.P. (1998). Endothelial alterations in hypercholesterolemia: more than simply vasodilator dysfunction. J. Cardiovasc. Pharmacol. *32* (Suppl. 3), S48–S53.

Ueno, S., Sano, A., Kotani, K., Kondoh, K., and Kakimoto, Y. (1992). Distribution of free methylarginines in rat tissues and in the bovine brain. J. Neurochem. *59*, 2012–2016.

Vallance, P., and Leiper, J. (2002). Blocking NO synthesis: how, where and why? Nat. Rev. Drug Discov. *1*, 939–950.

Vallance, P., and Leiper, J. (2004). Cardiovascular biology of the asymmetric dimethylarginine:dimethylarginine dimethylaminohydrolase pathway. Arterioscler. Thromb. Vasc. Biol. *24*, 1023–1030.

Vallance, P., Bush, H.D., Mok, B.J., Hurtado-Guerrero, R., Gill, H., Rossiter, S., Wilden, J.D., and Caddick, S. (2005). Inhibition of dimethylarginine dimethylaminohydrolase (DDAH) and arginine deiminase (ADI) by pentafluorophenyl (PFP) sulfonates. Chem. Commun. 5563–5565.

Accession Numbers

Coordinates have been deposited in the Protein Data Bank with accession codes 2CI1, 2CI3, 2CI4, 2CI5, 2CI6, 2CI7, and 2C6Z.

Chapter B - Membrane Proteins

B.1 Introduction to Membrane Proteins

Living organisms, from simple species to complex multicellular systems are formed from cells. Membranes, envelopes built up by lipids, achieve separation from the environment and distinguish between the inside and outside of the cell. Lipid molecules are hydrophobic molecules with polar head groups and long non-polar hydrocarbon tails. Depending on the molecules and the lipid composition, they build different high molecular assemblies. Monolayers of lipids form ball shaped micelles, whereas the membranes of the cells are generated from bilyers of lipids. Although there are proteins that enhance the diffusion capability of water through the membranes, lipid bilayers generally allow water and hydrophobic substances to diffuse freely into the cell. In contrast, polar substances are blocked and cannot pass the membrane and therefore need special “holes” to cross this barrier. The membranes of prokaryotes comprise one or two bilayers, which are separated by the periplasm. Higher organisms, like eukaryotes contain a cell wall and in addition, several organelles inside the cell, which are also enclosed by membranes, for example the nuclei, mitochondria, or the endoplasmatic reticulum.

Interaction with the environment or other compartments inside the cell across membranes is central to cell physiology, including uptake of nutrients, elimination of waste products, energy generation, signaling and secretion of peptides. Similarly to the chemical reactions that take place inside the cell, these processes across membranes require the presence of proteins. Located close to or inside the lipid bilayer, these proteins are called membrane proteins.

B.2 Membrane Protein Classification

Analysis of genomic data revealed that 30% of all proteins encoded on the genome of an organism interact with the lipid bilayers and are therefore called membrane proteins. Classification of these proteins is based on the different types of interaction with the membrane. In general, there are four subgroups: integral proteins which span the lipid layer several times, monotopic proteins that dip into the membrane but do not cross it, proteins anchored with one transmembrane helix and proteins that are anchored by post translational modifications with lipid molecules or glycosyl-phosphatidylinositol anchors (Figure 6). Depending on the secondary structure elements, which are used to span the membrane, the integral proteins are divided in α-helical or β-barrel transmembrane proteins. The latter are called porins and cross the membranes with an even number of β-strands generating a β-barrel, which creates an aqueous pore in the membrane. Members of the α-helical subgroup are for example: bacteriorhodopsin, a light driven proton pump that generates the energy for survival of Halobacterium salinarum (Luecke *et al.*, 1999); G-protein coupled receptors, which mediate signaling across the membrane (Lundstrom, 2005); K^+ or Cl^- channels (Zhou *et al.*, 2001, Dutzler *et al.*, 2003), which selectively allow transport of these ions through the membrane and photosystem I and II (Loll *et al.*, 2005, Ferreira *et al.*, 2004, Jordan *et al.*, 2001, Fromme *et al.*, 2001), the bases of photosynthesis and therefore life as we know it. Examples of β-barrel proteins are the Omp family of *Escherichia coli* which includes proteases like OmpT (Vandeputte-Rutten *et al.*, 2001) or anchoring proteins like OmpA (Pautsch & Schulz, 1998), unspecific pores like MspA from Mycobacterium smegmatis (Faller *et al.*, 2004), specific transport pores like LamB and ScrY (Buchanan, 1999) and active transporters like FepA, FecA, FhuA and BtuB (Locher *et al.*, 1998, Ferguson *et al.*, 2002, Chimento *et al.*, 2003, Buchanan, 1999), which use the proton gradient from the inner membrane to transport iron or vitamin B12 across the outer membrane. The location of the two subtypes of integral membrane proteins is strictly conserved, α-helical proteins are found exclusively in the inner membrane whereas β-barrels are located in the outer membrane of gram positive prokaryotes. Monotopic membrane proteins, including prostaglandin synthase (Picot *et al.*, 1994) and oxidosqualene cyclase (Thoma *et al.*, 2004), dip into only one leaflet of the membrane and import or extrude their hydrophilic substrates or products directly from or into the membrane.

Although membrane proteins represent a significant proportion of the genome, the structural data available is extremely limited. Only 106 different membrane proteins have been characterized structurally whereas the number of all entries in the Protein Data Bank exceeds 40000. The low proportion of membrane proteins reflects the low amount of expressed and purified material generally obtained and their amphipathic surface, which requires detergents to maintain the proteins in solution. A prerequisite of the structural determination of proteins by X-ray is the availability of protein crystals, where a small unit repeats several ten thousand times in all three dimensions to form a single crystal. This formation requires repetitive and unique protein-protein interactions between the molecules. For integral membrane proteins this area of interaction is rather limited, due to the fact that most of the protein surface is covered by detergent molecules and therefore not accessible for these specific interactions.

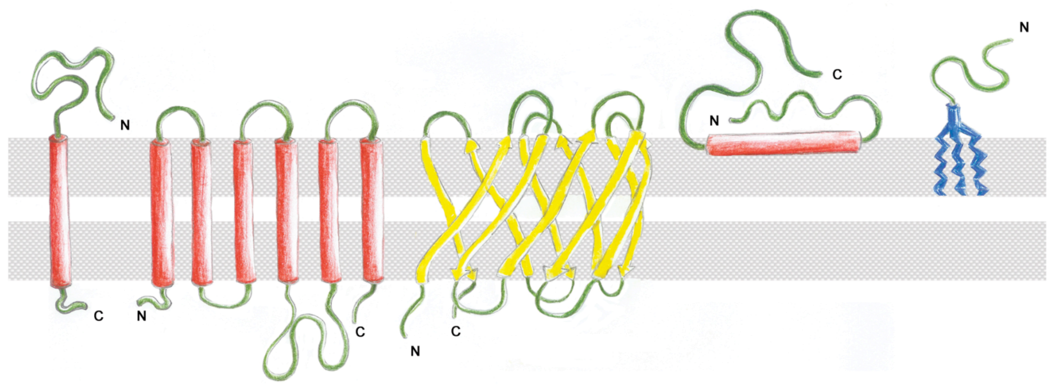

Figure 6: Schematic representations of different membrane protein classes. From left to the right, proteins that are anchored with one transmembrane helix, integral membrane proteins where α-helices cross the membrane or β-strands forming a β-barrel, monotopic proteins that dip into one leaflet of the lipid bilayer but do not cross them and a protein that is bound to the membrane via a GPI anchor.

B.3 Transport Proteins

The uptake of molecules by a cell requires the involvement of a transmembrane transport system that enables the entry of the substrate into the cytoplasm of the cell. The different transport systems identified so far, are classified into five groups that are categorized according to the transport process and their energy-coupling source.

Channels and pores build up group one of the transporter classification system (TC, http://www.tcdb.org/). They are passive transporters that facilitate the energy independent diffusion of small solutes across the membrane. They allow the diffusion of solutes across the membrane until equal concentration of the substrate is established. The outer membrane protein MsbA of *mycobacterium smegmatis,* consisting of two consecutive β-barrels, belongs to this family (Niederweis, 2003, Faller *et al.*, 2004). It enables the passive transport of small molecules across the membrane of the bacterium. Channels that show higher specificity towards the transported molecule are e.g. the glycerol facilitator GlpF of *E.coli* or the KirBac1.1 of *Burkholderia pseudomallei*. The selectivity of these systems allows the passive and specific diffusion of glycerol or potassium across the membrane thereby excluding the transport of water molecules or sodium ions, respectively (Fu *et al.*, 2000, Kuo *et al.*, 2003, Enkvetchakul *et al.*, 2004). Aquaporins, mediating the fast exchange of water molecules in and out of the cell are also members of this class of transport systems (Heymann & Engel, 2000).

Electrochemical potential driven transporters, also called secondary carrier-type facilitators, belong to the second category. The mediated transport event is coupled either to the proton motif force (pmf) or sodium motif force (smf). These transporters are further subdivided as uniporters, proteins that facilitate the transport of a single species either by diffusion or in a membrane potential-dependent driven process if the solute is charged, symporters that couple the transport of two or more species in the same direction and antiporters which transfer two or more solutes in opposite direction across the membrane. Prominent examples of this family are the lactose permease of

E.coli that couples the uptake of H^+ and β-galactosides (Abramson *et al.*, 2003) or the resistance-nondulation-cell division (RND) family that use the pmf to catalyze the substrate efflux in a H^+ antiport mechanism (Sobczak & Lolkema, 2005b).

Primary active transporters form the third group of the transport classification system. They utilize either chemical energy or light to generate a solute or ion-concentration gradient. The chemical energy transport can be P-P hydrolysis dependent, decarboxylation dependent, methyltransfer-dependent or oxidoreduction dependent. A member of this class of transporter is the ABC transporter superfamily, which uses the chemical energy stored in the phosphate bonds of ATP. This family will be discussed in more detailed in the next section. Other members are bacteriorhodopsin or the oxaloacetate decarboxylase Na^+ pump of *Klebsiellia pneumoniae,* which use light to generate a proton gradient or a decarboxylation reaction to create a sodium ion gradient, respectively.

Proteins belonging to the group of translocation transport systems make up the fourth category. This group of transporters differs from all the others since it is the only family, which modifies the substrate while it is transported. The translocators so far characterized are found exclusively in bacteria and belong to the phosphoenolpyruvate: sugar phosphotransferase systems. Here, the extracellular sugar is phosphorylated during transport and ends up as phosphate-sugar solute in the cytoplasm.

Systems, which enable an electron flow across the membranes, belong to the last and fifth group of the TC system. The electrons, which are generated in the cytosol, have to cross the membrane to reach their acceptor molecules. Well-characterized members are the DsbB and DsbD oxidoreducatase systems of *E.coli*. The DsbB pathway transfers electrons from periplasmic dithiol proteins to quinones in the bacterial membrane. The reduced quinones further reduce cytoplasmic acceptors like O_2, fumarate or nitrate. Transport of electrons from the cytosol to the periplasmic space is mediated by the DsbD system, where the electrons are transported from the cytosolic protein thioredoxin to the acceptor protein DsbC in the periplasm. The transport of electrons is achieved by disulfide shuffling between the membrane protein DsbD and the protein DsbC in the periplasm, which in the end reduces the target proteins (Raina & Missiakas, 1997, Stirnimann *et al.*, 2006).

B.4 Detergents

The complex nature of the lipid bilayer hinders many studies and experiments involving membrane proteins. It is, therefore, desirable to transfer the protein of interest to a more tractable environment. The new environment must stabilize the protein during its removal from the lipid bilayer, conserve its function and accommodate for the hydrophobic and hydrophilic nature of membrane proteins. The amphipathic structure of membrane proteins requires buffer systems that prevent interactions of the hydrophobic transmembrane segment with the surrounding polar water molecules. This border must be occupied by molecules that can interact with both the hydrophobic segments and the hydrophilic water and, thus, these molecules should also exhibit an amphipathic nature. Such molecules are detergents. Like fatty acids, they consists of a polar head group and a hydrophobic tail and are therefore capable of masking the hydrophobic surface of the membrane protein, which is normally embedded in the lipid bilayer, and simultaneously interact with the polar water molecules. These interactions prevent aggregation of the membrane protein molecules and keep them in solution.

Detergents are grouped into 4 different classes according to their chemical nature (Figure 7). Ionic detergents contain a positive or negative charge depending on their polar head group. In addition they contain a hydrophobic backbone that can be either linear or more complex e.g. steroid derivatives (Figure 7 a and b). The bile salts have a polar and an apolar face and therefore, they rather form kidney shaped aggregates unlike the spherical micelles formed by other detergents (Seddon *et al.*, 2004). Non-ionic detergents contain hydrophilic polar head groups of polyoxyethylene or glycosidic nature and a linear hydrophobic non-polar alkyl chain (Figure 7 c). Zwitterionic detergents have polar head groups with opposite charges, thus creating an overall electro neutral molecule and a linear or bulky hydrophobic alkyl chain (Figure 7 d).

The nature of the detergent molecule has tremendous effects on protein stability, activity and crystallization behavior. In general, the non-ionic detergents are considered as mild detergents, disrupting lipid-lipid interactions but not protein-protein interactions and are therefore most suitable for solubilizing a protein whilst retaining its function. Nevertheless, smaller alkyl chains of the non-ionic detergents often deactivate or denature the protein. Ionic detergents with linear alkyl chains for example SDS are harsh and often denature the proteins whereas the ionic bile salts are considered to be mild detergents (Seddon *et al.*, 2004). Despite this generalization, individual membrane proteins behave differently with the various classes of detergents and the best detergent suitable for each experiment must be determined experimentally.

Each detergent is characterized by several physical properties including the critical micelle concentration (CMC), the critical micelle temperature (CMT), the Kraft point, the cloud point and the aggregation number. The CMC is the maximal monomer concentration of the detergent in solution. Above this concentration the detergent forms micelles: globular aggregates of monomers. In these micelles the polar head groups of the detergent cover the majority of the surface and the hydrophobic tails of the detergent are on the inside. At low temperatures the detergents are mostly insoluble in water and form crystalline material. Upon raising the temperature, more and more monomers are dissolved in the solvent and at a certain temperature the concentration of monomers is high enough to form micelles. This temperature is called CMT. The Kraft point is the temperature in the phase diagram at which monomers, crystalline material and micelles coexist in equilibrium and it is, for most detergents, identical to the CMT. Solvents that contain non-ionic detergents undergo an additional phenomenon at the cloud point. Above this temperature the solution separates into a detergent rich phase and a detergent depleted phase. Finally, the aggregation number is defined as the number of monomers of detergent that are present in the formed micelle.

Parameters like the concentration of the detergent, the length of the alkyl chain, the ionic strength, the temperature, the pH value and the nature of the salt present influence the behavior and properties of the detergent in the solution. Ionic detergents show low aggregation numbers that depend on the ionic strength and composition of the solution whereas the aggregation numbers of non-ionic detergents are less influenced by the ionic strength.

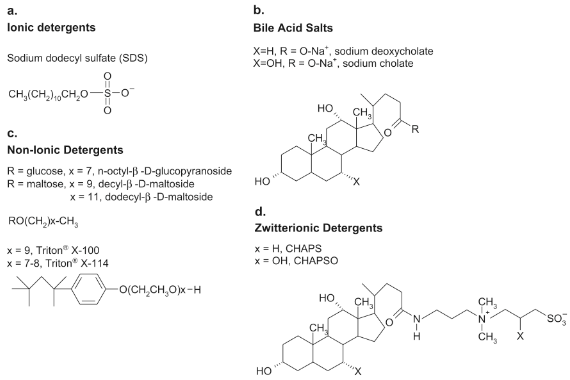

Figure 7: Examples of different detergents.

Taken from Seddon *et al.* (Seddon *et al.*, 2004).

B.5 Crystallization of Membrane Proteins

A prerequisite of a successful structure determination by X-ray is the availability of protein crystals. The formation of well-ordered crystals requires intermolecular interactions of a protein molecule to an adjacent one. These interactions must be specific and strong enough to build a three-dimensional ordered crystal. Unlike soluble proteins, which can use their entire surface to interact with the adjacent molecule, membrane proteins have only a limited hydrophilic surface area. One type of crystals can be obtained by adding lipids to the purified protein detergent mixture and removing the detergent. The so formed 2D crystals are in principle reconstituted biomembranes and are used with great success in electron microscopy (Stahlberg *et al.*, 2001). In crystals used for X-ray structure determination of membrane proteins the molecules arrange themselves to adopt one of the following types of crystal systems (Ostermeier & Michel, 1997). Type I crystals contain hydrophobic and detergent mediated contacts that align the individual protein molecules into two dimensional arrays. These 2D layers are then stacked on top of each other by polar contacts mediated by the hydrophilic parts of the protein. Due to this packing, these crystals often show anisotropic diffraction behavior. Figure 8 shows examples of membrane proteins crystallized in the different types of lattices.

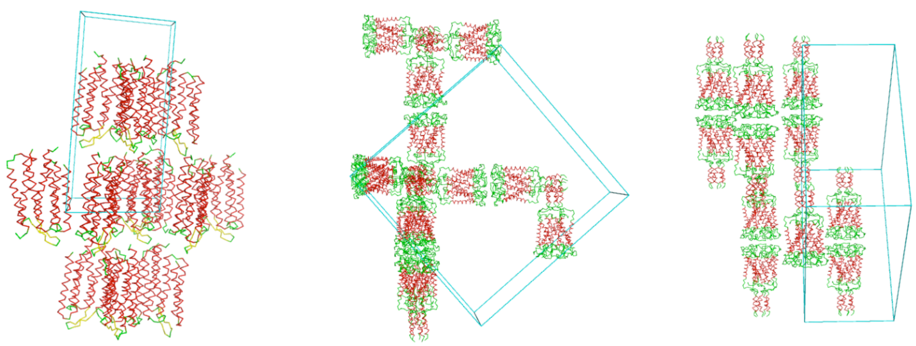

Figure 8: Crystal types of membrane proteins. Left image: Type I crystal packing of bacteriorhodopsin (PDB 1C3W). The red membrane embedded parts form a 2D lattice that is connected by polar interactions of the hydrophilic part (green and yellow). Center and right images: Crystal packing of a mechanosensitive channel (PDB 1MSL) type II packing. The red membrane spanning parts of the protein are not involved in crystals contacts. The green hydrophilic parts of the protein exclusively form the intermolecular interactions. The images are correlated by a 90° rotation.

Crystals of bacteriorhodopsin (Luecke *et al.*, 1999), which crystallized with Type I packing, were obtained by crystallization in lipidic cubic phases, a method explored by Landau & Rosenbusch (Landau & Rosenbusch, 1996). Type II crystals build the 3D lattice exclusively by hydrophilic contacts between their surface exposed loops and residues. Crystals of Type II are more common and are found for example in the crystal structure of the mechanosensitive channel (Chang *et al.*, 1998).

Several problems arise during the attempts to crystallize a membrane protein: (i) the stability of the membrane protein in the solubilized state, (ii) removal of structurally important lipids, (iii) the presence of the micelle around the hydrophobic belt of the protein, and (iv) the limited availability of hydrophilic regions suitable for crystal contacts. The problem of stability or loss of function was reported for the protein P-glycoprotein, which upon extraction from its natural environment exhibited a reduced or complete loss of function that could only be restored by reconstitution into a lipid bilayer or upon addition of lipids to the detergent protein mixture (Callaghan *et al.*, 1997). An effect of lipids on the crystallization behavior of several proteins has been reported, e.g. for LacY, cytochrome *b6f* and bacteriorhodopsin (Guan *et al.*, 2006, Zhang *et al.*, 2003, Belrhali *et al.*, 1999). The choice of the detergent is crucial, since the belt of detergent molecules must allow the close proximity of the hydrophilic parts. Larger detergents may interfere with the packing arrangement but detergents with a shorter alkyl chain are harsher to the protein and might denature it partly.

One way to increase the possibility of crystallizing a membrane protein is to increase the hydrophilic surface. This can be achieved by selecting a homologous protein containing longer surface exposed loops or by crystallizing of membrane protein in complex with a soluble protein. Crystallization of membrane proteins with their naturally occurring soluble interaction partners would therefore be desirable. However, these complexes should exhibit high affinity and homogeneity, the binding protein must be known and should be easily obtainable and show low internal flexibility (Ostermeier & Iwata, 2003). Since most of the naturally occurring complexes do not fulfill these requirements, Ostermeier et al (Ostermeier *et al.*, 1995) used synthetic complexes to increase the chances to obtain a crystal structure of the membrane protein of interest. Antibodies or fragments of antibodies have been used to determine crystal structures of soluble proteins, membrane proteins and DNA complexes (Kovari *et al.*, 1995). Co-crystallization of integral membrane proteins with antibodies was first employed by the laboratory of H. Michel (Ostermeier *et al.*, 1995, Iwata *et al.*, 1995, Hunte & Michel, 2002, Hunte *et al.*, 2000). In their experiments they used monoclonal Fv fragments. Proteolyic cleavage of an antibody generates two domains: the Fab and the Fc region. Fab fragments contain variable and constant domains of both heavy and light chains (Figure 9). The smallest fragment, Fv, consists of the two variable domains of the light and heavy chains. The structure of the CLC channel of *E.coli* (Dutzler *et al.*, 2003) solved to a resolution of 2.5 Å, gives some insights into the packing of an antibody membrane protein complex (Figure 10). The Fab fragments that bind the channel at the periplasmic side exclusively form the crystallographic contacts.

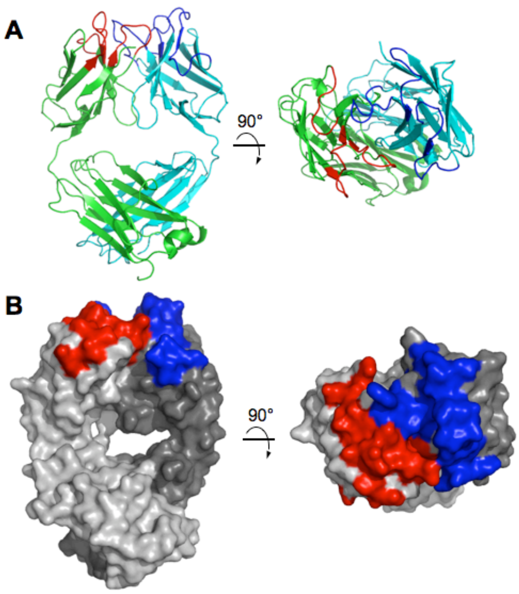

Figure 9: Position of CDR in a Fab molecule. A: The green cartoon represents the two Ig domains of the light chain with the CDRs marked in red. The constant and variable domains both consist of an Ig fold: the light chain is shown in green and the heavy chain in light blue. The CDRs are marked in red and blue respectively. The right picture is rotated by 90° around the central hinge region. B: Surface representation of the Fab f3p4 with the same orientation as in A.

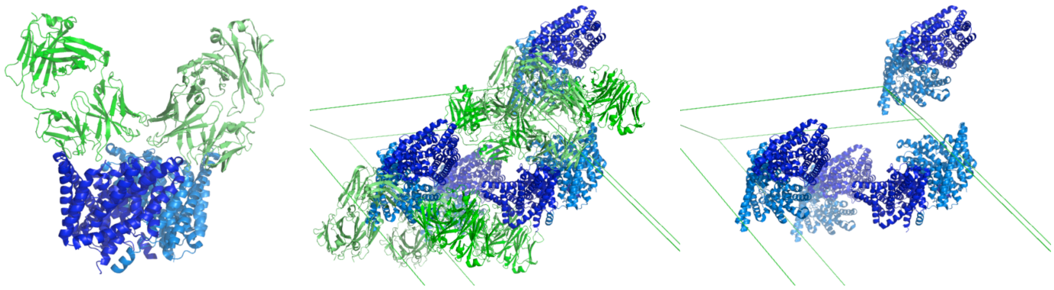

Figure 10: Structure and packing of CLC channel of *E.coli*. The dimer of the membrane protein and the two bound Fab fragments are shown in blue and green respectively. The Fab fragment binds to the periplasmic side of the transporter in a 1:1 stoichiometry. The center and right images illustrate the packing of the crystal. The CLC dimer is solely connected to the Fab fragment (center) visualized upon deletion of the Fab protein (right). No crystallographic contacts of adjacent dimer molecules are seen. PDB entry: 1OTS.

Designed ankyrin repeat proteins (DARPins), recently developed by Binz and co-workers (Binz *et al.*, 2004), are an additional possibility for complex generation. The generated combinatorial libraries can be used to probe for binding against different target molecules using ribosome or phage display. The modular architecture of the DARPins is built from 33-amino acids repeats consisting of a β-turn, two antiparallel α-helices followed by a loop connecting to the β-turn of the next repeat. The inner repeats (typically 2-3) are flanked by an N- and C-terminal capping repeat to shield the hydrophobic core from the solvent (Figure 11). The six randomized positions (X_2, X_3, X_5, X_{13}, X_{14}, and X_{33}) in the internal repeats generate a binding groove that is located in stable secondary structural elements. The highly diverse libraries contain synthetic binders with binding affinities that are comparable to antibody complexes (Huber et al., 2007, Kohl et al., 2003).

Complexes of DARPins with various binding proteins have been reported and structurally characterized. These complexes include structurally unaltered target molecules, e.g. maltose binding protein, and target molecules that were inhibited upon formation of the DARPin complexes, e.g. APH (Kohl et al., 2005, Amstutz et al., 2005) or Caspase-2 (Schweizer et al., 2007). A complex of DARPin molecules with the membrane protein AcrB, which is the most effective multi drug resistance protein of E.coli, was reported by Sennhauser et al. (Sennhauser et al., 2006). AcrB consist of three identical protein chains that build an asymmetric trimer of 340 kDa. Two DARPin molecules bind to the periplasmatic side of AcrB, thereby inhibiting the function as shown in in vivo experiments Figure 12

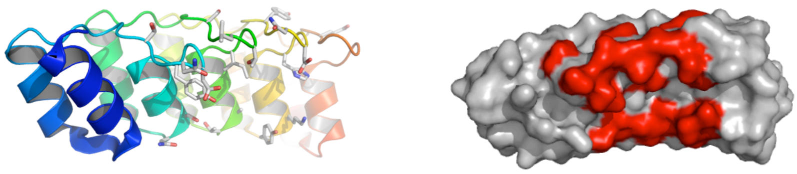

Figure 11: Structure of a designed Ankyrin molecule. Structure of a DARPin molecule (Kohl *et al.*, 2003). The variable positions are shown in stick representation. The blue N-terminal and red C-terminal capping repeats are constant to all library members. A surface representation with the variable positions shown in red is drawn on the right side.

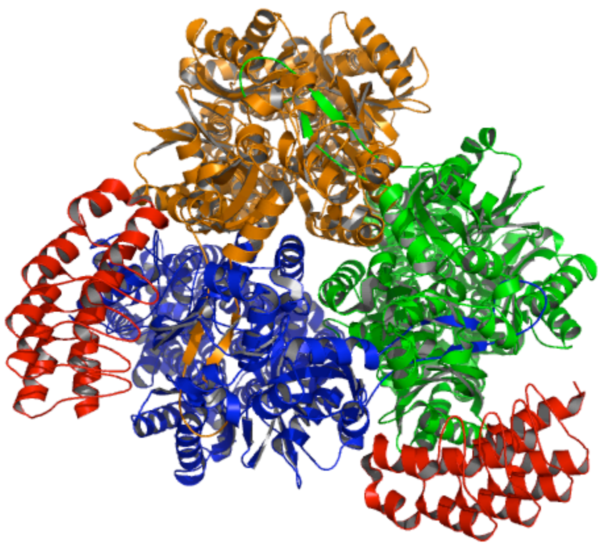

Figure 12: AcrB-DARPin complex. View along the pseudo three-fold axis of AcrB in the direction of the cytosol. The two DARPin molecules are shown in red and the three chains of AcrB are shown in blue, green and orange**.**

Chapter C - ABC Transporters

C.1 Introduction to ABC Transporters

ABC (ATP-binding-cassette) transporters form one of the largest protein families with more than 2000 different protein sequences. They are present in all kingdoms of life ranging from simple prokaryotes to eukaryotes to mammals. The importance of this family is shown by the fact that almost 5 % of the genome of *Escherichia coli* encodes for ABC transporters (Higgins & Linton, 2004, Linton & Higgins, 1998), which may reflect the great ability of this organism to survive in different environments. The human genome encodes for at least 48 ABC transporters which transport substances out of the cell, but are also involved in intracellular transport processes among different compartments. Eukaryotic ABC transporters are classified in 7 classes, ABCA, ABCB, ABCC, ABCD, ABCE, ABCF and ABCG based on gene structure (half or full transporter), domain arrangement and sequence homology in the nucleotide binding (NBD) and transmembrane domains (TMD). It should be noted that ABCE and ABCF do not contain membrane domains, but these proteins share homology in the NBD with other ABC proteins and are therefore categorized in this super family of proteins. Most of the ABC transporters are involved in genetic diseases (Igarashi *et al.*, 2004) and indeed it is hard to study any physiological system in depth without identifying the role for an ABC transporter (Higgins & Linton, 2004). Some known diseases, which are coupled to a malfunction of an ABC transporter, are listed in Table 1.

Most of the ABC proteins are active transporters utilizing the stored energy in ATP to translocate specific substances across the cell membranes, but some are involved in activities other than transport, e.g., regulation or targeting of other non ABC proteins. Bacterial ABC transporters can be classified as in- or exporters whereas their mammalian counterparts are in general export systems. Imported substances can be small ions, for example iron or nickel which are transported with the help of chelators such as siderophores, vitamins, for example vitamin B12, which serve as a cofactor for a number of enzymes, or all kinds of nutrients such as amino acids or sugars.

Table 1: Human diseases associated with ABC transporter

Transporter	Function	Disease
ABCB1 (MDR1) ABCC1 (MRP1) ABCG2 (MXR)	Drug resistance	Cancer
ABCC7 (CFTR)	Chloride channel	Cystic fibrosis
ABCA4 (ABCR)	Rod photoreceptor retinoid transport	Stargard disease and age-related macular degeneration
ABCA1 (ABC1)	Cholestorol and phospholipids transport	Tangier disease and familial HDL deficiency
ABCB11 (SPGP,Bsep) ABCB4 (MDR2)	Bile-salt transport	Progressive familial intrahepatic cholestasis
ABCC2 (MRP2)	Bile acid transport	Dubin-Johnson syndrome
ABCC6 (MRP6)	Unknown	Pseudoxanthoma elasticum
ABCC8 (SUR1) ABCC9 (SUR2)	Targeting and regulation of Kir6.2	Persistent hypoglycemia of infancy
ABCB7 (ABC7)	Iron transport	Sideroblastic anemia and ataxia
ABCD1 (ALD)	Very long chain fatty acids transport	Adrenoleukodystrophy
ABCG5, ABCG8	Sterol transport	Sitosterolemia
ABCB2 (Tap1) ABCB3 (Tap2)	Peptide transport	Immune deficiency

Data taken from Gottesman & Ambudkar (Gottesman & Ambudkar, 2001) and Stefkova et al (Stefkova et al., 2004)

Almost all of these importers contain, beside the actual ABC system, an additional protein, which is necessary for the transport process. These proteins, located in the periplasm of the prokaryotes, bind the substrate and direct it to the transport system where transport across the inner membrane takes place. In contrast to the highly specific importers, the export systems contain proteins, which have broader substrate specificity and translocate a large number of substances across the membrane. Export systems contain members involved in cell membrane maintenance such as the *E.coli* MsbA that uses the energy of ATP to invert lipids in the two layers of the inner membrane. Other export systems are the well studied Phospho-glycoprotein (PGP), a multidrug export system in humans, which extrudes hydrophobic substances of the cytosol or membrane out of the cell and confers multidrug resistance to cancer cells, the multidrug resistance-associated proteins (MRP) or the transporter associated with antigen processing (TAP; (Abele & Tampe, 2004), a transport system that plays a major role in immune defense. TAP transports antigen peptides into the endoplasmatic reticulum, where the MHC type I complex binds these peptides and presents them after export to the cell surface to T-cells.

Some ABC systems are not involved in transport activity but rather ATP binding or ATP hydrolysis induces conformational changes that regulate other cellular processes, for example K^+ channel regulation of Kir6.2 via the sulfonylurea receptor Sur1 (Huopio *et al.*, 2002, Chan *et al.*, 2003) or Cl^- channels such as the cystic fibrosis transmembrane conductance regulator (CFTR). CFTR is itself a Cl^- ion channel, but regulates a number of other Cl^- channels and helps in combating bacterial infections in the lung (Ko & Pedersen, 2001). Other ABC transporter family members, which are not involved in transport and even lack the TMD are, for example the two DNA repair proteins MutS and Rad50 (Higgins & Linton, 2004).

C.1.1 Sequence Similarity

Considering all the different activities of ABC transporters, it is not surprising that the overall sequence similarity of these proteins is rather low. TMDs share no or only low sequence similarity among the different ABC proteins and species, although they all share the hydrophobic character necessary to span the membrane. In general, the polypeptides span the membrane twelve times although up to 20 transmembrane helices are observed (Locher *et al.*, 2002, Abele & Tampe, 2004). The low sequence identity in the TMDs might be explained by the different specificities of the transporter. The pore forming subunits of the transport proteins define the substrate specificity and therefore the sequence variability of these domains of the protein is explicable.

In contrast to the TMDs, the different NBDs show, independent of prokaryotic or eukaryotic origin, in- or exporter, or the nature of the transported solute, a high sequence similarity; and this domain groups the different proteins in one

ABC super family (Higgins *et al.*, 1986). Sequence motifs present in the ~ 200 amino acids of the NBDs of ABC transporter include the WalkerA, WalkerB, A-, H- and Q-loop and the ABC signature motif (Figure 13). Walker A (GXXGXGKST, red) and the less conserved Walker B (grey) motif coordinate the nucleotide via hydrogen-bond interactions. The amino acid composition of Walker B is composed of a stretch of four hydrophobic residues followed by an aspartate, which makes the nucleophilic attack to the substrate (hhhhDEPT; (Orelle *et al.*, 2003). In contrast to these motifs that are also present in other nucleotide utilizing enzymes, the ABC signature motif (LSGGQ, magenta) and the A-, H- and Q-loop (yellow, cyan, orange respectively) are unique to ABC transport proteins. They are involved in the active site formation and bind water molecules and the nucleotide. Although their precise function in the catalytic cycle is less clear, it is known that the ABC signature motif interacts with the nucleotide of the neighboring NBD (Locher *et al.*, 2002). The ABC signature motif is located ~ 30 amino acids before and the conserved histidine in the H-loop (cyan) is approximately 30 to 40 amino acids after the Walker B (grey) motif. The Q-loop (orange) determines the end of a β-sheet and is located around 40 amino acids after the Walker A (red) motif.

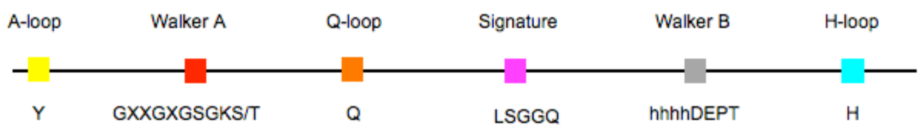

Figure 13: Schematic representation of conserved regions in the NBD of ABC transporters. From left to the right from (N to C terminus) the conserved regions A-loop, Walker A, Q-loop, the signature motive, Walker B and the H-loop are shown. The lower row marks the corresponding conserved amino acids. Amino acids are denoted by their single amino acid code, where X stands for every amino acid, S/T stands for either S or T and hhhh represents the hydrophobic segment included in the Walker B motif. Walker A and B (red and grey) are found also in other ATP utilizing proteins whereas the signature (magenta) and the two loops Q and H (orange and cyan) are found exclusively in ABC transporter systems.

C.1.2 Domain Arrangement of ABC Transporters

The classical ABC transporter consists of four domains; two α-helical hydrophobic transmembrane segments which generate the pathway for the substrate across the lipid bilayer; and two attached water-exposed cytosolic nucleotide binding domains (NBD) that mediate the conformational changes necessary during the transport process.

In prokaryotic systems these four domains can be encoded on two to four different polypeptides. The different domains are found arranged in a variety of ways and schematic representations and examples of such transporters are shown in Figure 14. The individual proteins build up the functional 2+2 units, either as homo- or heterodimers. An example of a transporter system containing four different proteins is the nickel uptake system Nik of *E.coli* that is formed by NikB, C, D and E. NikB and C are the two NBDs and NikD and E the hydrophobic pore forming membrane proteins. In addition, as with most of the uptake systems, this transport system contains an additional periplasmic protein NikA, which binds the substrate and locates it to the transporter (Figure 14A). The vitamin B12 uptake system from *E.coli* achieves the same transporter architecture with only two proteins. Two copies of the pore forming subunit BtuC and two copies of the NBD BtuD generate the functional transporter. BtuE, directing vitamin B12 to the transporter, is the corresponding periplasmic binding protein (PBP) of this uptake system. It binds the substrate after vitamin B12 is transported across the outer membrane (Figure 14B).

Gene fusions between the individual subunits are found in various transporter systems such as the arabinose uptake system AraGFH, where both NBDs are encoded on the protein AraG, the *E.coli* lipid flipase MsbA that contains a NBD and a TMD on one polypeptide as well as ferrichrome transporter FhuBC, which consist of two copies of the NDB FhuC with the two membrane spanning domains fused on the gene FhuB (Figure 14 C to E, respectively).

In general, uptake systems contain an additional protein, a binding protein that binds the substrate in the periplasm and locates it to the related transporter before permeation through the inner membrane takes place. Export systems achieve transport without additional proteins and therefore, ABC transporters, which have no assigned function, are generally categorized as importers or exporters depending on the occurrence of a periplasmic binding protein.

In contrast, eukaryotic ABC transporters generally consist of one polypeptide chain that encodes all four domains, but variations also occur here. A well studied example is the antigen presenting transport system Tap, where the formation of the active transporter is achieved by hetero dimerization of two different proteins, each of them containing one NBD and one TMD subunit (Figure 14F&G). The eukaryotic transporters, known to date, belong to the category of exporters, with no additional domains comparable to the bacterial PBP found within this class of proteins, although examples with additional domains at the intracellular side, for example the transporter CFTR, are described (Figure 14H).

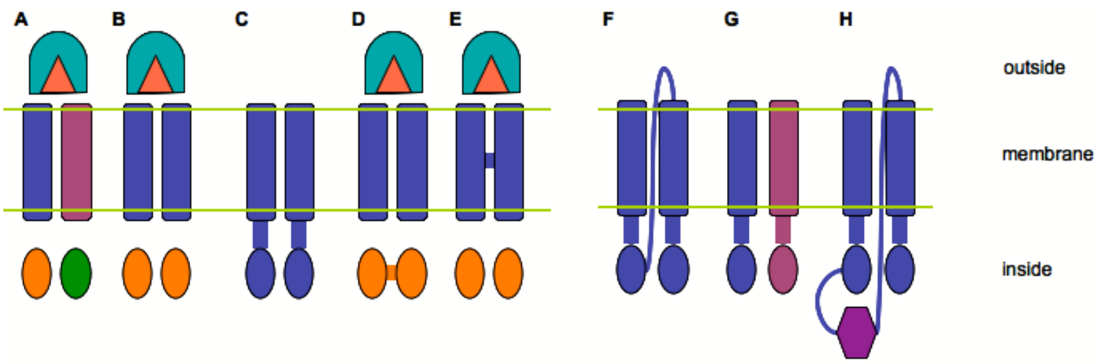

Figure 14: Different domain arrangements of ABC transporters. Gene fusion leading to multi domain proteins is a common feature in ABC transporter. Nucleotide binding domains are shown in circles and the transmembrane domains as elongated squares. Half circles with triangles represent additional binding proteins in the periplasm, which are present in the bacterial import systems. The domains are associated in various ways and examples of combinations that form the actual transporter are depicted in the scheme. **A:** four individual proteins, e.g. NikBCDE; **B:** two, but separated proteins, e.g. BtuCD; **C:** homodimeric assembly of fused TMD-NBD or NBD-TMD, e.g. MsbA; **D:** fused NBD-NBD and separated TMDs, e.g. AraG; **E:** fused TMD-TMD and separated NBDs, e.g. FhuB; **F:** all four domains in one chain, e.g. PGP; **G:** heterodimeric transporter with fused TMD-NBD, e.g. Tap system; **H**: full transporter with an additional domain in the cytoplasm, e.g. CFTR.

C.1.3 Structural Models

Although the existence of ABC proteins has been known for over 20 years, they were not structurally characterized until 2001, when Chang and coworkers published the first ABC transporter MsbA of *E.coli* (Chang & Roth, 2001). Until that point, structural characterization was based on electron microscopy data of PGP (Higgins *et al.*, 1997) and structures of ABC proteins that consist of the NBD but lack the transmembrane segments. Since then two more structures of complete ABC transporters have been published including the homologue of MsbA from *Vibrio cholare*, which contrasts with the closed conformation of to the *E.coli* protein (Chang, 2003), and the Vitamin B12 uptake transporter BtuCD from *E.coli* (Locher *et al.*, 2002). Although several groups reported biochemical studies of the cooperativity of the nucleotide binding domains (Karttunen *et al.*, 2001, Benabdelhak *et al.*, 2005, Moody *et al.*, 2002, Davidson *et al.*, 1996), the arrangement of the two NBDs with respect to each other was debated because of the different dimer assemblies present in the various solved structures of NBD dimers (Locher & Borths, 2004). In general a monomeric NBD consists of two lobes. Lobe I contains the Walker A (red), Walker B (grey), H-loop (cyan) motifs as well as the tyrosine in the A-loop (yellow) while the ABC transporter signature motif (magenta) is found in lobe II or the helical region. The Q-loop (orange) present in the linker region between the two sub domains is also known as γ-P_i linker.

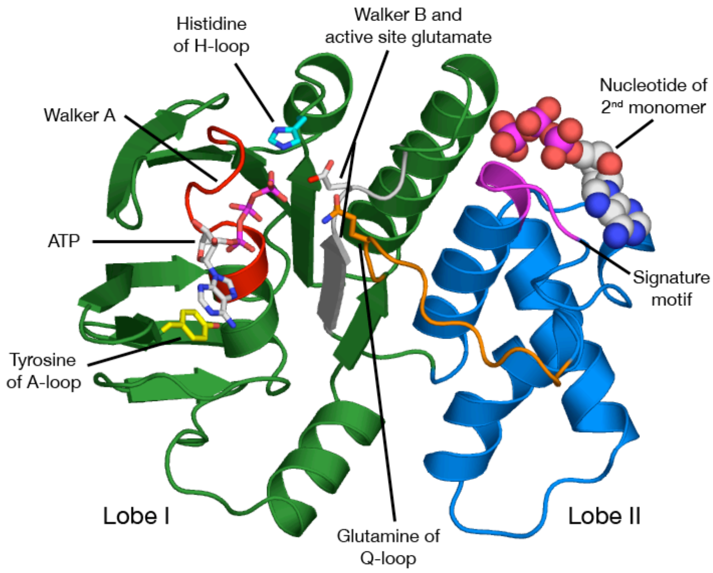

Figure 15: Cartoon representation of one NBD of Mj0796. The α/β domain (Lobe I) is shown in dark green and the purely helical Lobe II in blue. The conserved motifs Walker A (red), Q-loop (orange), signature motif (magenta) and Walker B (grey) are highlighted. The ATP molecule, H-loop (cyan), the active site glutamate (located close to the Walker B motif) and the glutamine providing the name for the Q loop are shown in sticks. The signature motif is in close contact to the nucleotide of the second monomer that is shown as spheres.

The dimer organization, first found in the ABC protein rad50, is now thought to be the biological arrangement of the NBDs and it is also observed in the BtuCD structure and in the dimer structure of MJ0796 E171Q mutant (Smith *et al.*, 2002). Figure 16 shows the two protomers in a head to tail arrangement, building up the two binding sites of the nucleotides at the interface between the molecules. The two signature motifs (magenta) generate the active site; they cap the triphosphates of the nucleotide (stick representation) bound to the Walker A (red) in the related subunit. The H-loop (cyan) makes contacts with the nucleotide and the neighboring protein subunit whereas the interaction of the Q-loop (orange) is restricted to the nucleotide in the same subunit. The tyrosine of the A-loop (yellow) interacts with the aromatic ring of the nucleotide. These tight interactions between the two nucleotides and the two subunits in an intra- and intermolecular way explain the reported cooperativity of the nucleotide binding domains. In addition, the two ATP molecules contribute approximately half of the buried area in the dimer interface, suggesting that ATP has an essential role in the formation of the dimer.

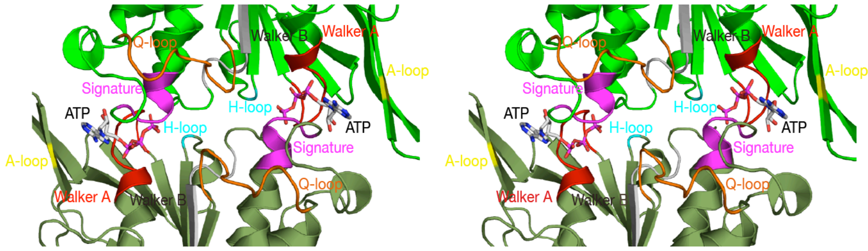

Figure 16: Stereo cartoon representation of Dimer interface of MJ0796. The dimer interface of MJ0796 E171Q mutant is shown (PDB code: 1L2T;(Smith *et al.*, 2002)). Mutation in the Walker B motif permitted binding but not the hydrolysis of ATP. The two NBD are shown in green and dark green and special sequence motifs are depicted as follows: red indicates the Walker A, grey color marks Walker B, magenta the signature motif for ABC transporter and A-, H- and Q- loop are shown in yellow, cyan and orange respectively. The two bound ATP molecules are shown in stick representation. See text for further explanations.

A detailed interaction scheme, involving the residues that bind the nucleotide in the dimeric structure of MJ0796 (Smith *et al.*, 2002), is shown in Figure 17. Each of the ATP molecules interacts with the Walker A (red) and B (grey) motifs, the A-, Q- and H- loop (yellow, orange and cyan respectively) of one NBD and with the signature motif (magenta) of the second NBD and therefore forming a sandwich like binding complex.

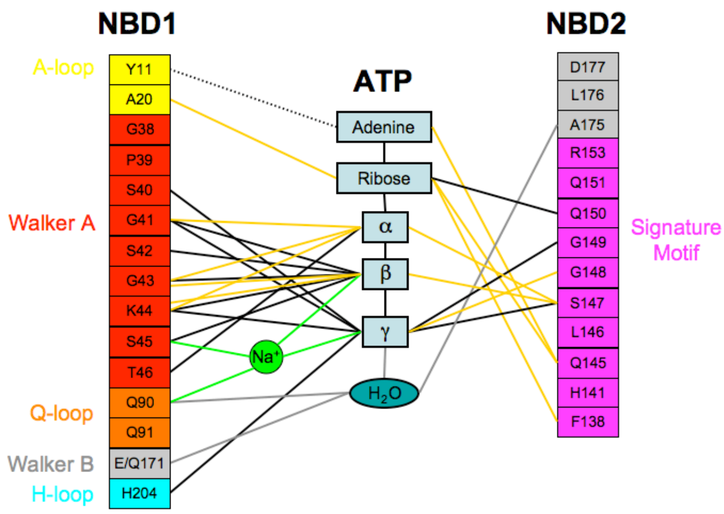

Figure 17: Nucleotide Interactions in the dimer interface of the two NBD subunits Schematic representation of the amino acids that interact with the ATP molecule. For clarity only one ATP molecule is shown. The NBD containing the Walker A and B motifs is shown on the right whereas the second NBD that contains the signature motif is shown on the right side. Van der Waals contacts are shown as orange lines, the aromatic stacking interaction of the tyrosine of the A- loop and the adenine base is shown in black dashes and the polar interactions of the two NBDs to the nucleotide are shown in black lines. Interactions with the catalytic co-factor (Na^+) or the active site water molecule are shown in green and grey. Adapted from Smith et al. (Smith *et al.*, 2002).

The ATP is bound by several hydrophobic and polar interactions from both NBDs. The catalytic co-factor (here Na^+) is bound to residues S45 of the Walker A and Q90 of the Q- loop of one NBD and to the β- and γ-phosphate of the nucleotide. The active site water is held in position by polar interactions of Q90 and E/Q171 of one NBD, the γ-phosphate of the ATP as well as of A175 of the second NBD.

Although several structures of NBDs of ABC transporters are present in the literature (Lewis *et al.*, 2004, Chen *et al.*, 2003, Diederichs *et al.*, 2000, Hung *et al.*, 1998, Smith *et al.*, 2002) the number of structurally characterized transporters including the membrane spanning parts is low. Up to now, structures of the Vitamin B12 uptake system BtuCD of *E.coli*, the multiple drug transport protein Sav1866 of *Staphylococcus aureus*, the putative metal-chelate transporter HI1470/HI1471 of *Haemophilus influenzae* and the three structures of the lipid flipase MSBA of *E.coli, Salmonella typhimurium* and *Vibrio cholerae* are known (Locher *et al.*, 2002, Dawson & Locher, 2006, Pinkett *et al.*, 2007, Chang & Roth, 2001, Chang, 2003, Reyes & Chang, 2005). The three structures of MSBA differ substantially from the other reported structures, particularly in the domain arrangement of the two NBDs, which is contradictory to the reported cooperativity. The observed MsbA structures rather reflect crystallographic artifacts and do not show the true and active arrangement of the transport system. In addition, the authors declared substantial errors during structure determination and therefore all structures of MsbA have been retracted and are under revision (Chang *et al.*, 2006).

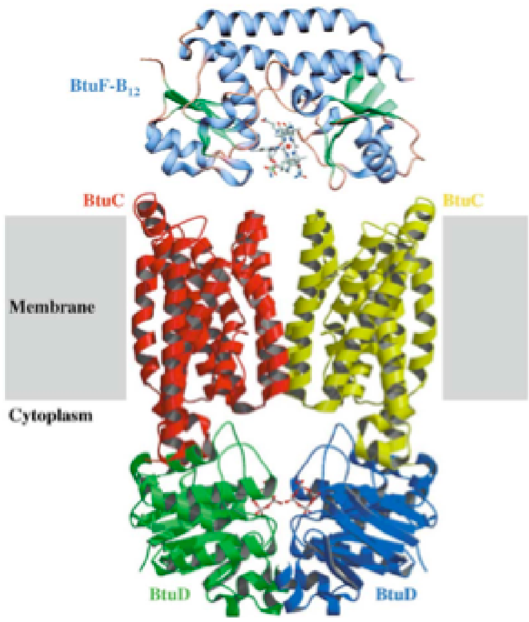

Figure 18: Structure of Vitamin B12 uptake system of *E.coli*. Cartoon representation of the BtuCD –BtuF-system. The structure was solved for the BtuCD complex that is comprised of two TMD subunits of BtuC (red and yellow) and two molecules of BtuD (green and blue). The periplasmatic protein BtuF-Vitamin B12 complex was placed manually in the complex and was not present in the structure (Taken from Locher; (Locher, 2004).

The structures of solved ABC transporters containing both membrane and soluble domains (Locher *et al.*, 2002, Dawson & Locher, 2006, Pinkett *et al.*, 2007) show similar features and have an arrangement supporting the biological findings. The structures show the two transmembrane domains surrounding the membrane translocation channel and the two attached energy proliferation domains. They are in close contact and share the widely accepted physiological interface that contains the conserved ATP-binding and ATP-hydrolysis motifs (Figure 18). In addition, the two structures of the BtuCD and HI1470/HI1471 show an alternating access to the inside or outside of the cell. A water filled cavity that is large enough to accommodate the substrates is accessible from the cytosol in the structure of HI1470/HI1471, whereas in the case of BtuCD the central cavity has access to the periplasm (Pinkett *et al.*, 2007). The opening and closing of the translocation pathway of the substrates can be explained by small rearrangements of the transmembrane domain triggered by movements in the NBDs.

C.1.4 Transport Energy Proliferation and ATP-switch Model

Up to now, it is still unknown when and how many nucleotides are hydrolyzed in one transport cycle and several models are proposed and discussed over the last decades to explain how ABC transporters achieve transport across the membrane (Higgins & Linton, 2004, Sauna *et al.*, 2006, van Veen *et al.*, 2000, van der Does & Tampe, 2004). The processive clamp model favors the hydrolyzing event of both nucleotides followed by a dissociation of the NBD dimer before reloading with ATP (Smith *et al.*, 2002, Janas *et al.*, 2003). The actual power stroke that mediates the transport is either the binding of the two ATP molecules and the subsequent dimer formation or, upon ATP hydrolysis the electrostatic repulsion of bound ADP and P_i could generate the mechanical energy that is used to rearrange the connected TMD to their ground state. This model is supported by crystal structures, which always contain equally occupied nucleotide binding sites in the two subunits (Smith *et al.*, 2002, Janas *et al.*, 2003).

However, studies of the vanadate trapped P-gp (Sauna *et al.*, 2001, Sauna *et al.*, 2006) and biochemical studies on the *E.coli* maltose uptake system MalFGK2 (Sharma & Davidson, 2000) showed asymmetric occupancy of the two NBDs. Only one ATP molecule was present and the ATPase activity of the second NBD was inhibited. Therefore a second model, the alternating access model was proposed. Here, one ATP molecule is hydrolyzed, the products ADP and Pi are released from one NBD and these events trigger the transport of the substrate. A second ATP is subsequently hydrolyzed to resets the transporter to its ground state (Ambudkar *et al.*, 2006). However, it should be mentioned that in the state where one NBD is in an open and one in a closed state, the dimeric interface would be strongly destabilized and therefore such a form has not been observed in structures solved with detached NBD. It is possible that an attached TMD in a complete transporter could stabilize this state (van der Does & Tampe, 2004).

The ATP switch model proposes several communication events between the TMD and the NBD and changes in the NBD dimer interface (Higgins & Linton, 2004). The model is based on two principal conformations for the NBDs: formation of a closed dimer upon binding of ATP at the dimer interface and dissociation to an open dimer after hydrolysis of ATP and release of ADP and P_i. Regulation of the transport and nonspecific ATP hydrolysis without transport of a substrate is regulated by signals of the TMD. The conformational changes also result in different affinities of the nucleotides and the substrate. Binding of the substrate to its high affinity site is the first step of the transport cycle, resulting in an increased affinity for ATP of the two nucleotide binding sites. In a second step ATP binds to each of the two the NBDs, which leads to a closed dimer between the two domains. This rearrangement of the NBD is communicated to the TMD enabling them to adopt a different conformation with a low affinity binding conformation for the substrate. The substrate then leaves the transporter and the hydrolysis of the two ATP molecules and the release of ADP and P_i finishes the transport cycle.

The proposed models of transport are predominantly based on well characterized export proteins. Except for the maltose uptake system MalGFK, in general, the importers are not very well characterized biochemically (van der Does & Tampe, 2004). Nevertheless, the recent structures of $ModB_2C_2A$ (Hollenstein *et al.*, 2007), Sav1866 (Dawson & Locher, 2006) and HI1470/HI1471 (Pinkett *et al.*, 2007) provided insight into a possible transport mechanism. Binding of two ATP molecules in the NBD dimer rearranges the attached coupled helices in the TMD, leading to a conformational shift of the inward-facing to the outward-facing conformation. Substrates bound to exporter could then be released to the outside of the cell. In contrast, in this conformation, the import proteins are accessible to their substrates opening a translocation pathway. Upon ATP hydrolysis, the NBDs and TMDs are rearranged to adopt their ground states and the opening of the transport channel releases the substrate to the cytoplasm (Hollenstein *et al.*, 2007).

C.2 Results for ABC Transporters

The aim of this work was the structural and biochemical characterization of an ABC transporter. To date three X-ray structures of complete transport systems from bacteria are known and published. In addition, there are low-resolution EM images for example, of the human phospho-glycoprotein that allow a rough estimation of the architecture of these proteins.

C.2.1 Eukaryotic ABC Transporter

The mouse Bile salt export pump (mBsep) is a eukaryotic member of the ABC transporter superfamily. It is located at the canalicular (apical) membrane side of hepatocytes in the liver and is involved in the transcellular transport of bile acids. The uptake of bile acids in hepatocytes is mediated by various non ABC transporters, such as the sodium-taurocholate co-transporting polypeptide (Ntcp) and the sodium independent organic anion transporters (OATPs) (Stieger & Meier, 1998). The extrusion of the bile acids is mediated by the multi drug resistance proteins (MDR) and the bile salt export pump, which are all members of the ABC transporter family.

MBsep or ABCB11 is a large protein of 1321 amino acids with a molecular weight of 147 kDa and belongs to the subfamily B of the ABC transporters (Dean *et al.*, 2001); http://www.ncbi.nlm.nih.gov/books/). Analysis of the secondary structure predictions of the protein indicates the existence of 9 -11 transmembrane helices, depending on the algorithm used (Figure 19). The protein contains two nucleotide binding domains (NBDs) located between and after the two transmembrane domains (TMDs); therefore it is thought that this gene is the product of a gene duplication of a former half transporter ancestor. Since both NBDs must be cytosolic, the number of transmembrane helices between the two NBDs must be even. The predicted helix in the region of residue number 900 (Figure 19) is likely to form two helices, and the helices at residues 1000 and 1030 are true helices. 12 transmembrane helices can therefore be assumed for the structure of mBsep.

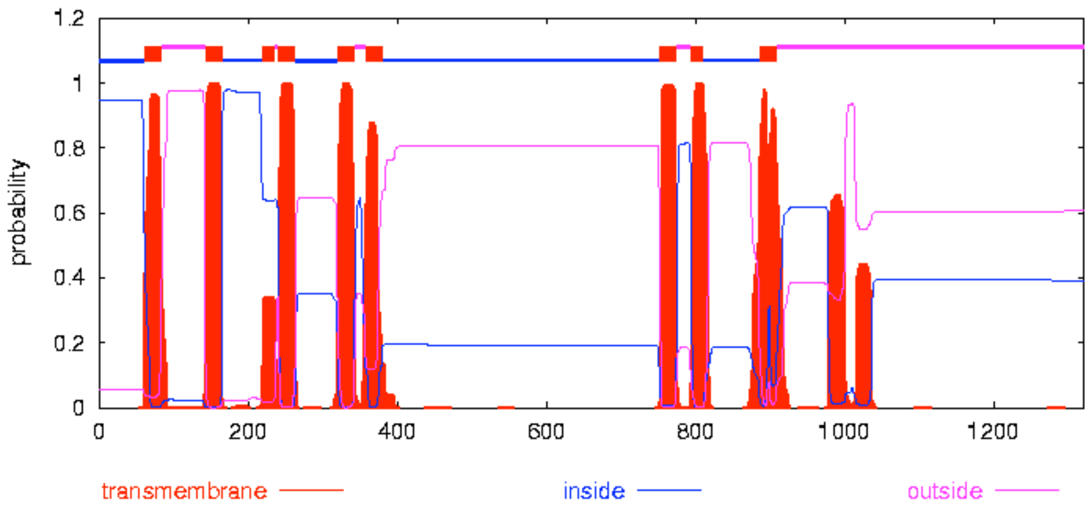

Figure 19: Prediction of transmembrane helices of mBsep. The TMHMM server predicts 9 helices, whereas helices 3 and helices 10 to 12 are not clearly defined (http://www.cbs.dtu.dk/services/TMHMM/). To locate both NBDs in the cytoplasm, the number of transmembrane segments in the second half of the molecule must be even.

C.2.1.1 Protein Expression in Insect Cells

MBsep was expressed in Sf9 insect cells in collaboration with B. Stieger of the University Hospital of Zurich. Infected Sf9 cells were harvested and resuspended in lysis buffer (50 mM Tris, 50 mM mannitol, 2 mM EGTA, pH 7.0) containing 1 ug/ml antipain and leupeptin and 1 mM PMSF. The cells were homogenized and intact cells and nuclei removed by centrifugation (10 min, 5000*g, 4°C). The membranes were collected (1 h, 100000*g, 4°C) and suspended in buffer (5 mM HEPES, 25 mM sucrose, 50 mM KNO_3, 6.25 mM $Mg(NO_3)_2$, pH 7.4), homogenized and stored at -80°C. The amount of protein produced was estimated by silver stained SDS PAGE and Western blot (Figure 20). Expression of mBsep could be detected on the Western blot but the silver stained SDS-PAGE showed no difference in pattern and intensity of the protein bands upon infection. This indicated a very low expression level of mBsep.

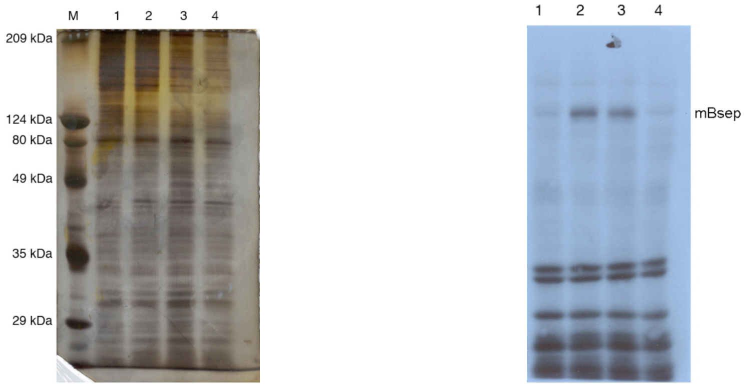

Figure 20: Silver stained 7.5% SDS-PAGE and the corresponding Western blot. M: molecular weight marker; 1: Sf9 cells; 2: mBsep; 3: mBsep control expression; 4: Sf9 cells infected with wild type virus. Total amount of protein was 10 ug /lane on the silver stained SDS-PAGE and 100 ug /lane on the Western blot. The upper band in lane 2 and 3 on the Western blot shows the expressed mBsep band with to an apparent molecular weight of 130 kDa**.**

C.2.1.2 Solubilization of mBsep

200 µl of membrane vesicles were incubated with 1 ml solubilization buffer (10 mM HEPES, 50 mM KNO_3, 25 mM sucrose, 0.02 % (w/v) NaN_3, 6 mM $MgSO_4$, pH 7.4) containing N-octyl-β-D-glucoside, CHAPS or N-dodecyl-β-D-maltoside at concentrations of 5, 5 or 5-100xCMC, respectively. The solution was stirred for 1 hour at 4°C. The samples were centrifuged and the protein concentration in the supernatant was determined (Advanced Protein Assay, Cytoskeleton). 900 µl of the supernatant was incubated at 4°C overnight with Ni-NTA material (Quiagen). The beads were collected and washed three times with solubilization buffer containing the detergent at a concentration of 3x CMC. The protein was eluted with 250 mM imidazole in the detergent/ solubilization buffer. The elutate was applied to SDS-PAGE and analyzed by Coomassie staining or Western blot (Figure 21).

None of the above detergents were suitable to achieve the solubilization and enrichment of mBsep from the Sf9 cell vesicles. Western blot analysis of the eluted fractions, visualizing the N-terminal His-tag of the protein and ATPase activity experiments showed similar results.

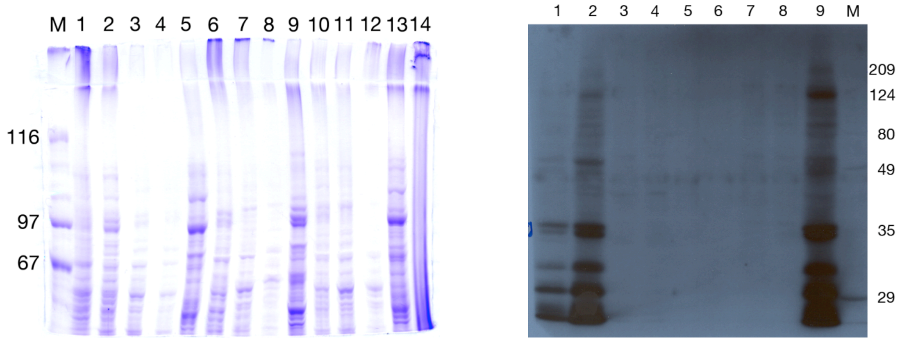

Figure 21: Solubilization of mBsep. Right: all lanes were loaded with 10 µg of total protein. M: molecular weight marker; 1: vesicles; 2: DDM extract; 3: DDM flow through; 4: DDM wash I; 5: DDM elution; 6: β-OG extract; 7: β-OG flow through; 8: β-OG wash I; 9: β-OG elution; 10: CHAPS extract; 11: CHAPS flow through; 12: CHAPS wash I; 13: CHAPS elution; 14: external marker. Left: 1: 10 µg vesicles; 2: pellet after solubilization; 3: supernatant after solubilization; 4: flow through; 5: wash I; 6: wash II; 7 elution I; 8: elution II; 9: 100 µg vesicles; M: molecular weight marker.

C.2.1.3 Cloning and Expression of mBsep in *Escherichia Coli*

The expression yield in Sf9 cells was far too low to conduct biochemical or structural studies of mBsep. Therefore, the following constructs were generated to facilitate the expression of mBsep in *E.coli*. Constructs of the full transporter with an additional cleavable N-terminal 6xHis-tag as well as constructs containing half of the transporter were generated. Full length and half transporters were PCR amplified using the appropriate primers and the mBsep pFASTBac HTb DNA as template. The PCR products were purified and ligated into the bacterial expression vectors pET20ch and pET20NHT. The two half transporter constructs comprised amino acids 1-755 or 756-1320 of the cDNA of mBsep. The cleavage site was chosen according to the start of the first transmembrane helix of the second transmembrane domain. The bacterial expression of mBsep constructs was carried out in *E.coli* BL21 (DE3), *E.coli* BL21(DE3) CodonPlus RIL and *E.coli* BL21 (DE3) plysS. The different *E.coli* strains were transformed with the plasmid of interest and plated on a LB agar plates containing the appropriate antibiotic. A single colony was used to inoculate a 3 ml preculture that was subsequently grown at 37 °C over night. 100 ml LB media was incubated with 1 ml of preculture and grown at 37°C to an OD_{600} of 0.6. Protein expression was started using 0.2 mM IPTG and the cells were grown at 18, 25 or 37 °C. The cells were harvested 1-8 hours after induction and a 1 ml sample was loaded on an

SDS-PAGE. The expression tests were repeated with 2YT media. Various expression tests with different strains, temperatures, media and IPTG concentrations showed no expression of the full or half transporters.

C.2.1.4 Cloning and Expression of mBsep in Mammalian Cells

Full-length mBsep was PCR amplified using the plasmid mBsep pFASTBac HTb as template. The resulting PCR product, containing the full-length mBsep, the N-terminal hexa His tag and the rTEV cleavage site, was cloned into the pEAK8 vector digested with *Hind III* and *Eco RI*.

The vector pEAK8-mBsepEGFP was produced by PCR amplification of full-length mBsep from the mBsep-peak8 vector including the N-terminal addition of the 6xHis-tag and cleavage site but lacking the 3` stop codon. The DNA of the enhanced green fluorescence protein was PCR amplified using the vector pEGFP-N1 as template. Both PCR products were sequentially cloned into the pEAK8 vector using *Hind III* and *Eco RI* or *Eco RI* and *NotI*, respectively. 0.5 x 10^6 293T cells were transformed with 10 µg of the appropriate vectors. After 7 h of incubation the cells were washed with 5 mL PBS buffer and grown in 8 ml media for an additional 20 h. The rate of transfection was checked by fluorescent microscopy with the control vector EGFP. The cells were harvested and loaded on an SDS Gel and the proteins were visualized with an anti-HIS antibody (Figure 22, left).

No signal was detected at a molecular mass around 120 kDa. Therefore the amount of DNA used for transfection was varied in the range 1-20 µg and the time of expression was varied from 1 to three days. A total amount of 150 µg of protein was applied to an SDS-PAGE and subsequently transferred to a cellulose membrane. The proteins were visualized with the primary antibody against mBsep and a secondary antibody, which was coupled to the enzyme horse radish peroxidase (Figure 22, right). Small amounts of mBsep could be detected upon transfection of the cells with 10 – 20 µg vector and an incubation time of 48 hours. Nevertheless the control of the mBsep expressed in Sf9 cells showed higher expression levels.

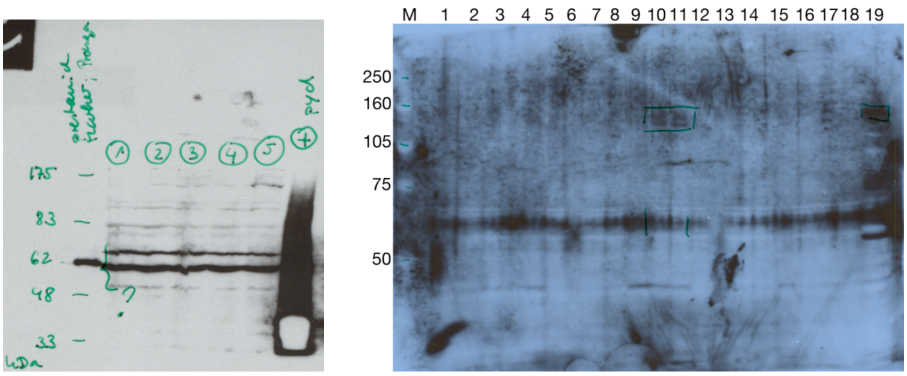

Figure 22: Expression tests of mBsep in mammalian cells. Left image: 293T cells transfected with 0,1,2,5 and 9 µg of mBsep-peak8 are shown in lanes 1-5, respectively. As a positive control, a vector containing PYD was used. The protein was detected using an antibody recognizing His-tagged proteins. Right image: 293T cells transfected for 24 h with 1,5,10 and 20 µg of mBsep-peak8 are shown in lanes 2-5. A GFP control plasmid was loaded on lane 1 and the cell without transfection in lane 6. Lanes 7-12 and 13-19 are the same experiment, except, with expression time of 48 and 72 h.

Although sufficient protein was produced to perform activity tests in membrane vesicles (Stieger *et al.*, 2000), the amount of the eukaryotic ABC transporter mBsep produced in all tested heterologous expression systems was barely detectable. Having available only small amounts of protein, for purification, structural studies, and biochemical characterization become impractical and, therefore, the work with mBsep was discontinued and the focus was shifted towards ABC transporters originating from different bacterial strains.

C.2.2 Prokaryotic ABC Transporters

The following section on prokaryotic ABC transporters describes the work and progress of the selection process, expression, purification and biochemical analysis of successfully selected prokaryotic ABC transporters.

C.2.2.1 Strategy

The main bottleneck to obtaining structural and functional information about membrane proteins is the availability of large amounts of the desired proteins. The heterologous expression and preparation large amounts of functional membrane proteins is influenced by various parameters, including protein stability, posttranslational modifications for example glycosylation or phosphorylation, lipid composition of the host, membrane thickness, toxicity and protease stability (Wagner *et al.*, 2006). Furthermore, the solubilization of the expressed proteins in an appropriate detergent is critical. Removal of the proteins from their natural lipid environment decreases the stability of the polypeptides during solubilization. This often results in nonhomogenity or precipitation of the sample mainly due to unfolding of the protein, depending on the detergent selected. However, in addition, to the mentioned problems during production and purification of the membrane protein, a further complication arises during crystallization. The dynamic nature of many transport proteins leads to multiple conformations, which, if present in the sample are unfavorable for the formation of well ordered crystals.

Working simultaneously with homologues of the protein of interest from different organisms is one strategy to overcome these problems. For example, instead working with a highly complex and unstable eukaryotic or human protein a bacterial homologue might be chosen. In cases where homologues an identical function are not present or accessible, homologues with different functions, but inheriting for example a conserved reaction mechanism, are an alternative to understand the protein of interest.

To gain insights into the general function and architecture we started with a selected set of 28 ABC transporters, including importers and exporters, transporters consisting of two or four polypeptide chains and proteins that originate in different bacteria. This should enhance the chances to obtain a transporter suitable for structure determination.

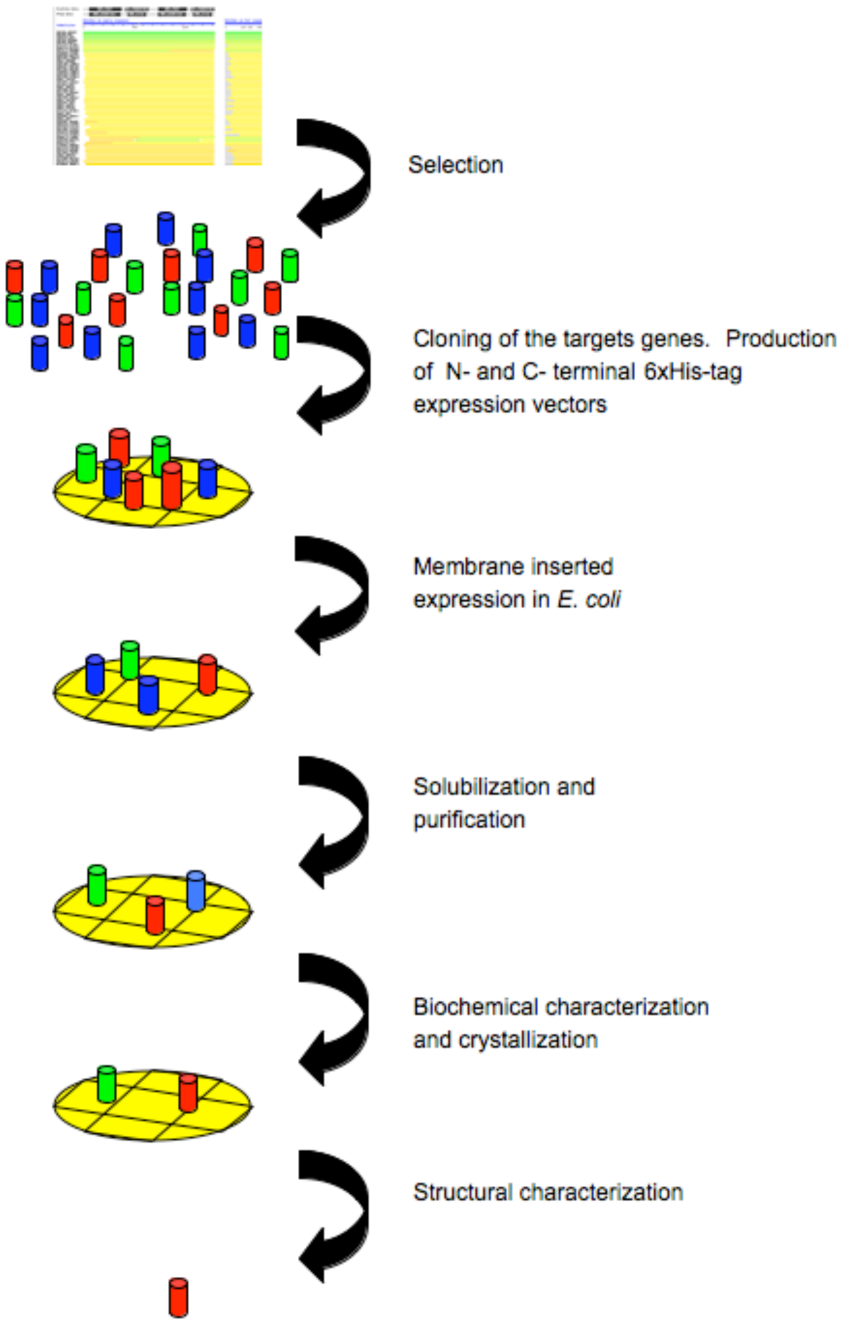

Figure 23: General strategy applied in the work on ABC transporters. The proteins of interest were selected and cloned into bacterial expression vectors. Different *E.coli* strains were transformed with these vectors and expression of the proteins was examined. Proteins with highest expression yields were subjected to purification, characterization and crystallization.

C.2.2.2 Selection

Since ABC transporters are occurring in all kingdoms of life and transport a wide range of different substrates, it is not surprising that the number sequences found in the ABC transporter superfamily is over 2000. Conserved sequence features that are located in the cytosolic domain of the transporters, harboring the ATP binding site, identify sequences of ABC transporters. In contrast, the membrane associated regions of the transporters are less conserved. These domains are responsible for the recognition of the substrate and therefore may show variations that allow transport of the different compounds across the membrane.

Starting with the protein sequence of mBsep, a blast search was performed to find possible homologues of this transporter in bacteria, but the enormous numbers of different ABC transporters and the high sequence identity present in the NBDs, resulted in too many diverse candidates. The annotated SwissProt database (Bairoch2004) was therefore used to identify possible targets for the project. A subset of the ABC transporters was selected according to two criteria:

1. Proteins should originate from thermophilic organisms.

These bacteria or archeabacteria live in an environment that reaches temperatures up to 100°C. Normally functioning at these high temperatures, it is clear that these proteins should exhibit higher stability and lower flexibility at room temperature. In particular, the potential increase in the chance of trapping the transporter in a distinct state of its catalytic cycle makes these proteins attractive targets for crystallographic studies.

2. Simplified cloning and expression

Simplified cloning and expression procedures could be achieved using those transporters that contain a fusion of TMD and NBD. This restriction reduces the number of different clones that have to be produced from each individual gene. In addition such constructs enhance the chance of obtaining a protein sample that contains the full transporter complex.

Despite these ideal's, however, other considerations influenced the selection of the target and various genes were chosen that do not fulfill the above mentioned criteria. For example it was argued to be advantageous to choose proteins originating from *E.coli* as they might express better in *E.coli* than proteins from other organisms. It was also decided to choose transport systems with known function so that material obtained could be characterized biochemically and functionally. Unfortunately, the annotated sequences mostly originate from *E.coli* and the majority of *E.coli* ABC transporters consist of an NDB and TMD on separate polypeptide chains. Nevertheless six of these transport systems were included in the pool of targets.

Bacteria were selected and their genomes scanned for the presence of NBD domains containing the signature motif of the ABC transporter family. Hits were analyzed with TMHMM (http://cbs.dtu.dk/services/TMHMM) to verify the presence of transmembrane helices. In the case of proteins that did not contain transmembrane helices, the two neighboring ORFs situated directly up and downstream on the genome were analyzed for the presence of transmembrane helices, and, if helices were detected, included in the work. The proteins and the corresponding open reading frames that were selected and used in this work are summarized in Table 2.

Figure 24 shows the alignment of a subset of the selected transporters and the similarity of the NBDs; especially highly conserved are the two motifs walker A and B. These two features are also present in other nucleotide binding proteins, however, the ABC transporters contain an additional, so called, signature motif, that distinguish them from other classes of nucleotide binding protein.

Table 2: Selected Proteins

Name	Species	Domain arrangement[1]	Substrate	Accession number[2]
TM0043	*Thermotoga maritima*	h	Unknown	Q9WXQ0
TM0287	*Thermotoga maritima*	h/m	Unknown	Q9WYC3
TM0288	*Thermotoga maritima*	h/m	Unknown	Q9WYC4
TM1256	*Thermotoga maritima*	h	Unknown	Q9X0Y3
TM1310	*Thermotoga maritima*	h	Unknown	Q9X136
TM1319	*Thermotoga maritima*	h	Unknown	Q9X144
TM1328	*Thermotoga maritima*	h	Unknown	Q9X151
HP0600	*Helicobacter pylori*	h	Multi drug resistance	O25322
HP1082	*Helicobacter pylori*	h	Lipids	O25714
PH1735	*Pyrococcus horikoshii*	h/m	Unknown	O59382
PH 1736	*Pyrococcus horikoshii*	h/m	Unknown	O59381
YFIB	*Bacillus subtilis*	h/m	Unknown	P54718
YFIC	*Bacillus subtilis*	h/m	Unknown	P54719
YHEH	*Bacillus subtilis*	h/m	Unknown	O07549
YHEI	*Bacillus subtilis*	h/m	Unknown	O07550
YKNU	*Bacillus subtilis*	h/m	Unknown	O31707
YKNV	*Bacillus subtilis*	h/m	Unknown	O3170
YDDA	*Escherichia coli*	h	Unknown	P31826
YOJI	*Escherichia coli*	h		P33941
YWJA	*Bacillus subtilis*	h	Unknown	P45861
MacB	*Escherichia coli*	h	macroliden	P75831
YBIZ	*Escherichia coli*	h	Unknown	Q8XED0
araG/H	*Escherichia coli*	m	L-arabinose	P0AAF3* P0AE26
glnP/Q	*Escherichia coli*	m	L-glutamate	P0AEQ6 P10346*
mglA/C	*Escherichia coli*	m	Methyl-galactoside	P0AAG8* P23200
tauB/C	*Escherichia coli*	m	taurine	Q47538* Q47539
xylG/H	*Escherichia coli*	m	xylose	P37388* P0AGI4
znuB/C	*Escherichia coli*	m	zinc	P39832 P52648*
mBsep	*Mus musculus*	f	Bile acids	Q9QY30

[1] f: full transporter, h: half transporter, m: two different polypeptides assemble the transporter, h/m two transporters neighbored on the genome that might form a heterodimer. [2]: (Bairoch *et al.*, 2005) UniProtKB/TrEMBL database, * denotes the ATP binding subunits of the transport systems.

```
BTUD          1 --------------------------------------------------MSIVMQLQDVAESTR--------L 16
mBsep_aa001 366 VIIAAMNIGNASSCLEIFSTGCSAASSIFQTIDRQPVMDCMSGDGYKLDRIKGEIEFHNVTFHYPSRPEV-KIL 438
mBsep_aa755 270 IAMSATAVGRTFSYTPSYAKAKISAARFFQLLDRKPPIDVYSGAGEKWDNFQGKIDFIDCKFTYPSRPDI-QVL 342
TM0288      303 SRQFTRPLNELSNQFNMIQMALASAERIFEILDLEEEKD--DPDAVELREVRGEIEFKNVWFSYDKKKPV---L 371
MJ0796        2 ------------------------------------------------------IKLKNVTKTYKMGEEIIYAL 21

BTUD         17 GPLSGEVRAGEILHLVGPNGAGKSTLLARMAGMTS-GKGSIQFAGQPLEAWSATKL---ALHRAYLSQQQTPPF 86
mBsep_aa001 439 NNLSMVIKPGETTAFVGSSGAGKSTALQLIQRFYDPCEGMVTLDGHDIRSLNIRWL----RDQIGIVEQEPVLF 508
mBsep_aa755 343 NGLSVSVDPGQTLAFVGSSGCGKSTSIQLLERFYDPDQGTVMIDGHDSKKVNVQFL----RSNIGIVSQEPVLF 412
TM0288      372 KDITFHIKPGQKVALVGPTGSGKTTIVNLLMRFYDVDRGQILVDGIDIRKIKRSSL----RSSIGIVLQDTILF 441
MJ0796       22 KNVNLNIKEGEFVSIMGPSGSGKSTMLNIIGCLDKPTEGEVYIDNIKTNDLDDDELTKIRRDKIGFVFQQFNLI 95

BTUD         87 AT-PVWHYLTLHQ--HDKTRTELLN----------DVAGALALDDKLGRSTNQLSGGEWQRVRLAAVVLQITPQ 147
mBsep_aa001 509 ST-TIAENIRLGR--EEATMEDIVQAAKDANAYNFIMALPQQFDTLVGEGGGQMSGGQKQRVAIARALIR---- 575
mBsep_aa755 413 DC-SIMDNIKYGDNTKEISVERAIAAAKQAQLHDFVMSLPEKYETNVGIQGSQLSRGEKQRIAIARAIVR---- 481
TM0288      442 ST-TVKENLKYGN--PGATDEEIKEAAKLTHSDHFIKHLPEGYETVLTDNGEDLSQGQRQLLAITRAFLA---- 508
MJ0796       96 PLLTALENVELPL--IFKYRGAMSGEERRKRALE-CLKMAELEERFANHKPNQLSGGQQQRVAIARALAN---- 162

BTUD        148 ANPAGQLLLLDEPMNSLDVAQQSALDKILSALCQQ-GLAIVMSSHDLNHTLRHAHRAWLLKGGKMLASGRREEV 220
mBsep_aa001 576 -KP--KILLLDMATSALDNESEAKVQGALNKIQH--GHTIISVAHRLS-TVRSADVIIGFEHGTAVERGTHEEL 643
mBsep_aa755 482 -DP--KILLLDEATSALDTESEKTVQLALDKARE--GRTCIVIAHRLS-TIQNSDIIAVMSQGVVIEKGTHKKL 549
TM0288      509 -NP--KILILDEATSNVDTKTEKSIQAAMWKLME--GKTSIIIAHRLN-TIKNADLIIVLRDGEIVEMGKHDEL 576
MJ0796      163 -NP--PIILADEPTGALDSKTGEKIMQLLKKLNEEDGKTVVVVTHDIN-VARFGERIIYLKDGEVEREEK---- 228

BTUD        221 LTPPNLAQAYGMNFRRLDIEGHRMLISTI--------------------------------------------- 249
mBsep_aa001 644 LERK------GVYFMLVTLQSQEDNTHKETGIKGKDTTEGDTPERTFSRGSYQDSLRASIRQRSKSQLSHLSHE 711
mBsep_aa755 550 MDQK------GAYYKLV--------------------------------------------------------- 560
TM0288      577 IQKR------GFYYELFTSQYG---------------------------------------------------- 592
MJ0796      229 --LR------GFDDR----------------------------------------------------------- 235
```

Figure 24: Alignment of NBD of 4 different ABC transporters. Sequences of BTUD, mBsep, TM0288 and MJ0769 are shown. Walker A and B motifs, Q-, D-, and H-loops and the signature motif are shown in red, gold, green, magenta, cyan and blue, The alignment was created with T-Coffee (Notredame *et al.*, 2000) and edited with JalView (Clamp *et al.*, 2004).

C.2.2.3 Cloning of Prokaryotic ABC Transporters

DNA of the selected ABC transporters was amplified by PCR using genomic DNA as the template and the product was subsequently cloned into an *E.coli* expression vector. The primer pairs were designed to match their theoretical melting temperature. According to the restriction enzyme used, 5-7 bases were added at the 5` end of each primer to enhance the digest activity. The PCR product obtained was purified on an agarose gel (Figure 25), digested and ligated into an expression vector, which was subsequently amplified and analyzed by restriction enzyme analysis and sequencing.

To reduce the number of PCR steps and to allow a fast transfer of the different constructs to different vectors, the following expression vectors were generated before the first cloning steps. Starting with the commercially available vectors pET20, pET21/28, pET28, and pTFT, three vector series were created containing an identical multiple cloning site (Table 3 to Table 5).

The majority of the vectors encode additional tags, which add the 6xHis-tag either at the N- or C- terminal end of the protein of interest. This tag enables the detection by His-tag specific antibodies and, furthermore, allows purification of the tagged protein by metal chelate chromatography; a robust and fast method to isolate the protein of interest.

Table 3: Modified pET20 Vectors

pET20-series, Ampicillin resistance	Properties
pET20b(+)	N-terminal *pelB* signal sequence, optional C-terminal 6xHis-tag, T7 promoter
pET20-no-Tag-no-pel	-
pET20-no-Tag-pel	*pelB* signal sequence for export to the periplasm
pET20-GST	N terminal fusion of the GST protein
pET20-GST-c-6xHis	N terminal fusion of the GST protein C-terminal 6xHis-tag
pET20-c-6xHis	C-terminal 6xHis-tag
pET20-c-6xHis-pel	*pelB* signal sequence for export to the periplasm, C-terminal 6xHis-tag
pET20-n-6xHis	N-terminal 6xHis-tag
pET20-c-6xHis-n-6xHis	C-terminal 6xHis-tag N-terminal 6xHis-tag

Table 4: Vectors used for co-expression of 2 proteins

pET21, pET28-series	Properties
pET21-c-6xHis	C-terminal 6xHis-tag Ampicillin resistance
pET28-c-6xHis	C-terminal 6xHis-tag Ampicillin resistance
pET21-n-6xHis	N-terminal 6xHis-tag Kanamycin resistance
pET28-n-6xHis	N-terminal 6xHis-tag Kanamycin resistance

Table 5: High and low copy plasmids

pET28/TFT-series	Properties
pET28-c-6xHis-new	C-terminal 6xHis-tag Kanamycin resistance
pET28-n-6xHis-new	N-terminal 6xHis-tag Kanamycin resistance
TFT-c-6xHis	N-terminal 6xHis-tag
Low copy plasmid	Ampicillin resistance
TFT-n-6xHis	N-terminal 6xHis-tag
Low copy plasmid	Ampicillin resistance

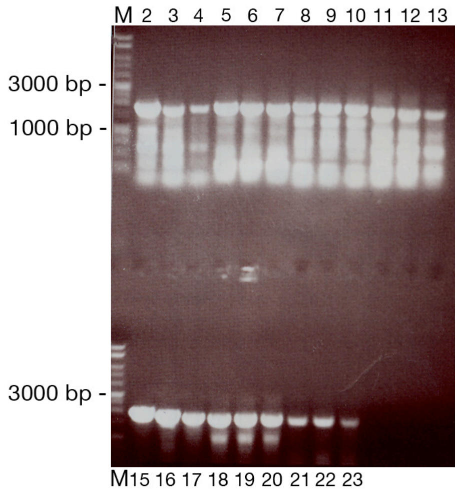

Figure 25: Resulting PCR products. 10 µg genomic DNA was amplified with the appropriate primer pairs. Lane 1: Marker DNA; lanes 2-4: TM0043; lanes 5-7: TM0287; lanes 8-10: TM0288; lanes 11-13:TM1256, lane 14: marker DNA; lanes 15-17:TM1310; lanes 18-20: TM1328; lanes 21-23: PH1736. Nearly all PCR showed additional band at molecular weights, which were smaller than the actual transporters ~1.5 kbp.

C.2.2.4 Expression of Prokaryotic ABC Transporters

The next section describes a general protocol that was used for the expression, purification and characterization of the selected ABC transporters mentioned in Table 2. More detailed protocols of individual transporters are given in the following chapter. A summary of all the expression tests with the various constructs and different expression cells can be found in the appendix in Table 8.

The plasmids encoding for the selected and successfully cloned ABC transporters were subjected to expression analysis in *E.coli*. Freshly transformed cells were grown over night in 2 ml pre-cultures, transferred to 100 ml liquid media and grown to an OD_{600} of ~ 0.5. After induction with IPTG samples were taken after 1, 3, 5 and 18 hours and subjected to SDS-Page analysis. In general, different types of media, different concentrations of IPTG and two induction temperatures were tested to obtain the highest expression levels for each transporter. Overexpression of the transporters could not be visualized on Coomassie stained gels, therefore, all expression tests were analyzed by Western blot. The protein was visualized using an antibody that recognizes N- or C-terminal His-tags (Quiagen) or an antibody specific to a C-terminal 6xHis-tag (Kaufmann *et al.*, 2002).

In general, it was observed that all transporters expressed poorly using the pET20b vectors. Transformation of the *E.coli* strains BL21(DE3), CD41 or CD43 (Miroux & Walker, 1996) with the generated plasmids lead to no or only very few colonies on a normal LB-Agar plate. After a supplement of 1% glucose to the LB-agar media more colonies were observed on the plates directly indicating toxicity in the cells upon induction of the heterologously expressed transport proteins. Therefore, all transporters were recloned into vectors based on the pET21 and pET28 series which have much greater control of the basal protein production. These vectors allow a stricter control of the expression levels before induction and are therefore more suitable for the production of proteins that are toxic to the host cell. In addition to the above-mentioned variations of the culture conditions, experiments with controlled pH during growth of the cells and nutrient feeding were carried out in a parallel fermenter.

C.2.2.5 Purification

The transporters that showed the highest expression levels in *E.coli* were further examined. The expression of each of the transporters was maximized by slight variations of protocols used during the expression tests. After determining the optimal expression parameters for each transporter, the protein was subjected to the purification process. After expression, the cells were pelleted by centrifugation and opened using a FrenchPress. The membrane fraction was collected and analyzed by SDS-PAGE and Western-blot. This allowed to find out whether the protein of interest had been correctly inserted in the membrane of *E.coli* or was disposed of in inclusion bodies. The membranes were subsequently solubilized using different buffers and detergents at various concentrations. The solubilized material was subjected to an ultracentrifugation step to remove the non-solubilized material and subsequently loaded onto metal chelate resin that had been previously equilibrated with nickel ions (chelating agarose Pharmacia; Ni-NTA Quiagen). After washing the column with increasing amounts of imidazole, the bound protein was eluted with a high concentration of imidazole. The eluted protein solution was analyzed by SDS-PAGE and size exclusion chromatography.

Size exclusion chromatography experiments were useful and necessary to determine the oligomeric state of the transporters after purification. Most of the transporters which showed high expression levels could be solubilized and purified by Ni-chelate chromatography. Unfortunately, however, the majority failed the size exclusion chromatography analysis either showing high amounts of aggregated protein or eluting with an estimated molecular weight that reflected rather a monomer than the proposed multimeric functional complex. In particular, the transporters that consist of two polypeptide chains, for example araG/H or tauB/C, were shown to be prone to fall apart with the two different polypeptide chains eluting at different volumes during the size exclusion chromatography run. In the cases mentioned above, where the transport proteins behaved unexpectedly during the size exclusion chromatography analysis, different pH values, salt concentrations and detergents were chosen and applied to the previous purifications steps to try to enhance the mono dispersity of the protein sample and increase the fraction of correctly formed complex.

C.2.3 Protocols for Specific Prokaryotic ABC Transporters

C.2.3.1 Proteins Originating from *Thermotoga maritima*

TM0287 and TM0288

The two ABC transporters TM0287 and TM0288 originate from the bacterium *Thermotoga maritima* strain MSB8. Both proteins contain a fusion of a transmembrane and a nucleotide-binding domain and belong to the so-called half transporters. TM0287 contains 577 amino acids with a molecular weight of 64360 Da. Prediction of the transmembrane helices indicated six clearly defined helices located between amino acids 7 and 297 and the cytosolic C-terminal NBD. Between helices one and two and helices five and six two shorter loops with a length of 30 and 14 amino acids respectively, are predicted to be located at the extracellular side of the membrane (Figure 26).

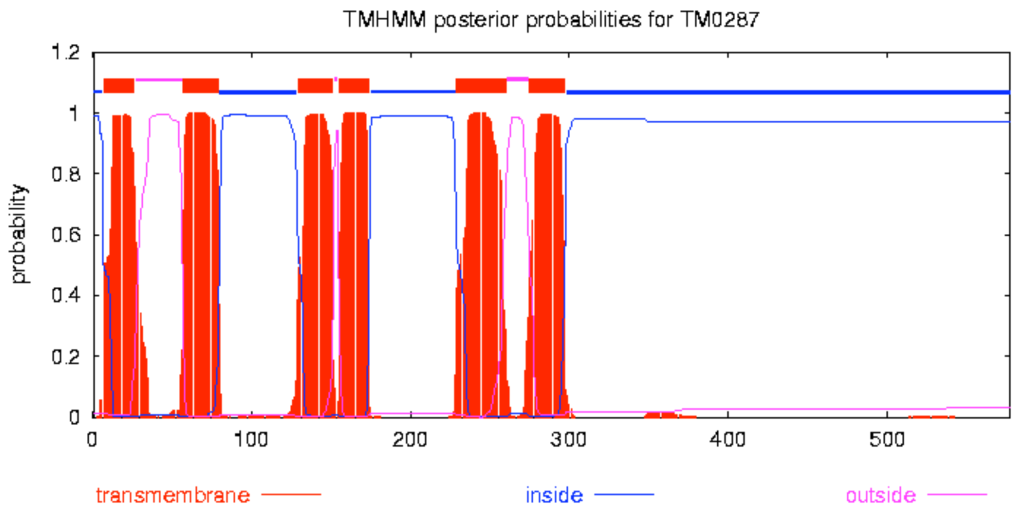

Figure 26: Prediction of the transmembrane helices in the ABC transporter TM0287. The red bars describe the localization of the transmembrane helices in the protein sequence. Two shorter loops outside of the cell are found between helices one and two or between helices five and six. (http://www.cbs.dtu.dk/services/TMHMM/).

TM0288 consist of 598 amino acids with a molecular weight of 67708 Da. The prediction of the transmembrane helices suggested 4 clearly defined transmembrane segments and one less well defined helix 5 (Figure 27). The prediction suggests the N-terminus is in the cytosol, but the 5 predicted helices place the C-terminal NBD domain incorrectly. It is obvious, therefore, that the broad peak in the prediction for helix 5 reflects two transmembrane helices rather than the one indicated by the algorithms. Splitting helix 5 into helices 5 and 6 would result in the correct cytosolic localization of the NBD. The protein contains a short external loop between helices 1 and 2 with a length of 23 amino acids.

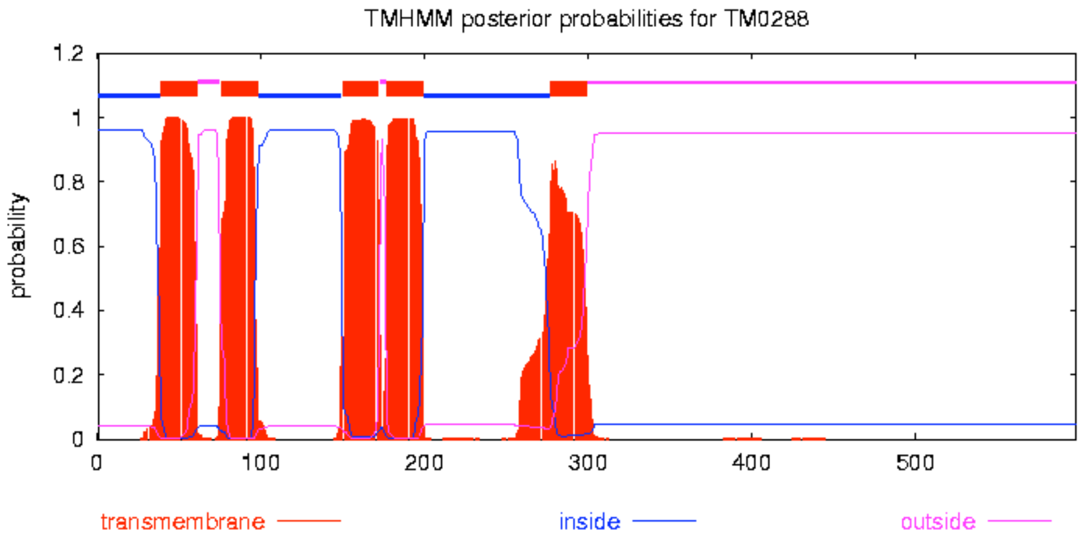

Figure 27: Prediction of the transmembrane helices in the ABC transporter TM0288. The red bars describe the localization of the transmembrane helices in the protein sequence. The broad red around amino acids 270-310 rather reflects two transmembrane helices, resulting in a total number of six helices and a correct localization of the C-terminal nucleotide-binding domain. Between helices one and two, a short external loop with a length of 23 amino acids is predicted. (http://www.cbs.dtu.dk/services/TMHMM/).

On the genome of *T. maritima* TM0287 is located between base pairs 303150-304883, whereas TM0288 is encoded by the base pairs 304876-306672. The two proteins share 8 bases encoding for the C-terminal aspartate, alanine and the stop codon of TM0287 or the starting methionine, proline and glutamate of TM0288 (Figure 28). Because of this close proximity it is not clear, if both genes encode for one hetero-dimeric transporter or if both proteins form homodimers, resulting in two individual ABC transporters. To address this question, the two proteins were expressed either alone or together in one host.

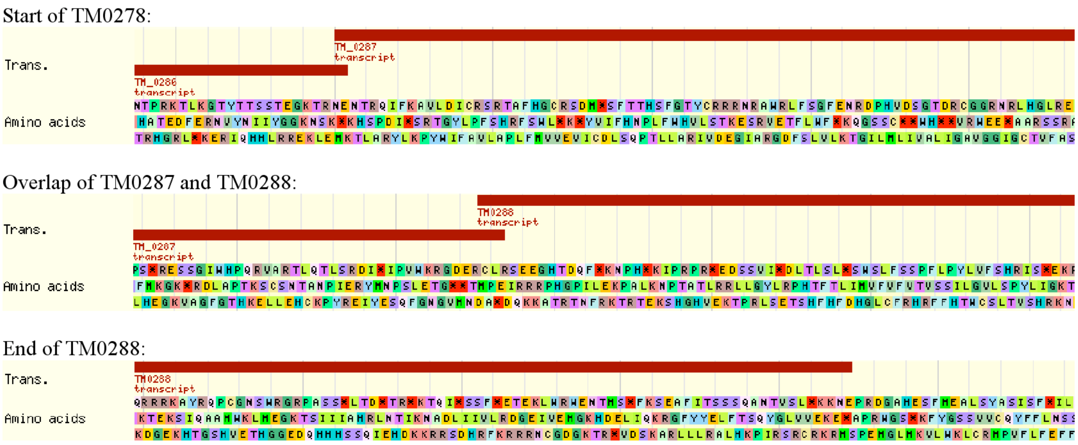

Figure 28: Gene organization of TM0287 and TM0288. The red bar in all three panels is the transcript of the DNA sequence. Upper panel shows the 5`end of TM0287, located at base pair 303150. The center panel shows the overlapping region of the two ABC transporters TM0287 and TM0288. Both genes make use of base pairs 304876 to 304883 either as the C-terminal end of TM0278 that is translated in frame 3 or as the N-terminal starting sequence in frame 2 used by TM0288. The lower panel shows the end of TM0288 encoding the stop codon at base pairs 306669-72.

Expression and Purification of TM0287 or TM0288

Expression levels of the transporters TM0287 and TM0288 were rather poor during the normal procedure for expression in shaker flasks. Therefore, a pH-controlled experiment in a parallel fermenter was carried out. The pH could be controlled externally with acid and bases (HCl and NaOH) or by the metabolites glucose and ammonia (acid/base). The latter allowed additional nutrient feeding during growth of the bacteria, but initial experiments showed that this method was prone to instability of the pH values due to the different growth behaviors of the bacteria. Therefore, all experiments were carried out with the HCl-NaOH combination to keep the pH at a constant value of 7.4.

A shaker flask containing 250 ml of 2YT media (16g/l bacto tryptone, 10g/l yeast extract, 5 g/l NaCl) and the appropriate antibiotic was inoculated with an overnight culture of BL21(DE3) containing the pET28-CH construct of TM0288, to give a starting OD_{600} of 0.1-0-15. The cells were grown at 37°C to an OD_{600} of 0.6-0.7, supplemented with 1% (w/v) glucose, induced with 0.2 mM IPTG and protein expression was carried out at 30°C for 3 h. The cells were harvested by centrifugation (30 min, 4000xg 4°C), flash frozen in liquid nitrogen and stored at -20°C.

Cells of a 3L culture were suspended in 60 ml buffer A (20 mM Tris, 20 mM NaCl, pH 7.5) containing 20 µl Dnase (30 mg/ml), 20 µl RNase (5 mg/ml), 1 mM PMSF and complete protease inhibitor cocktail (Roche). The cells were lyzed by 3 passes in the FrechPress, were centrifuged to remove soluble proteins and the resulting pellet was incubated for 1h at 4° with 110 ml buffer A containing 1 mM PMSF, complete Protease inhibitor cocktail (Roche) and 1% (w/v) DDM. Non-solubilized material was removed by ultra-centrifugation (100000 x g 4°C) and the supernatant was loaded on a 2 ml Chelating Sepharose column loaded with nickel (Pharmacia) that had been equilibrated in buffer A containing 0.03% DDM. Prior to elution with buffer A containing 250 mM imidazole and 0.03% DDM, unbound sample and impurities were removed by washing the column with 100 ml buffer A containing 0.03 % DDM and 50 ml buffer A containing 50 mM imidazole and 0.03 % DDM. The purification process is summarized in Figure 29. The protein TM0288 migrated with an apparent molecular weight of 65 kDa.

Elution 1-3 ml was concentrated to 180 µl and analyzed by size exclusion chromatography (Superdex 200 23 ml, Pharmacia) using an isocratic elution in buffer A supplemented with 0.03 % DDM. The protein eluted in a single peak at 10.6 ml corresponding to a size of ~130 kDa (Figure 30). Fractions containing pure protein were concentrated and stored at 4°C. Analysis of the sample after two weeks showed that the protein was stable at 4°C. No degradation or different elution properties of the size exclusion chromatography column fractions were observed.

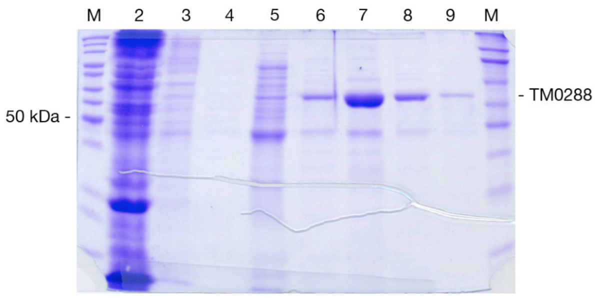

Figure 29: Purification of TM0288. The SDS-PAGE shows the enrichment of the transporter during purification. Lane 1+10: Marker proteins; lane 2: 2 µl of the membrane fraction; lane 3: 2 µl of the flowthrough of the Ni-column; lane 4: 5 µl of the wash without imidazole; lane 5: 30 µl of the wash with 50 mM imidazole, lane 6: 30 µl of elution 0-1 ml; lane 7: 10 µl of elution 1-3 ml; lane 8: 30 µl of elution 3-5 ml; lane 9: 30 µl of elution 5-7.

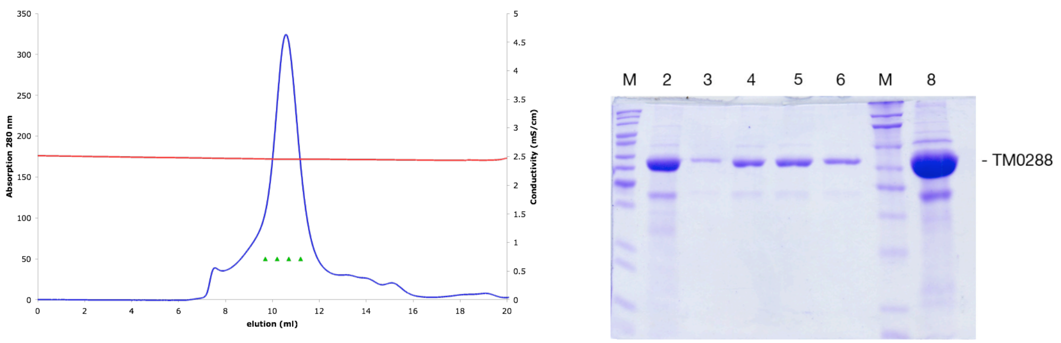

Figure 30: Size exclusion chromatography and corresponding SDS-PAGE of TM0288. The protein eluted in a sharp peak at an elution volume of 10.6 ml corresponding to a molecular weight of 130 kDa. In addition a small amount of aggregate is visible at 7.8 ml. SDS-PAGE: lane 1 +7: Marker proteins; lane 2: 2µl of the concentrated sample prior to size exclusion chromatography; lane 3-6: 10 µl of the fractions 9-12; lane 8: 2 µl of pooled and concentrated fractions; lane 9-10: 30 µl of the flowthrough from both concentrations steps. The arrow indicates TM0288.

To verify the presence of folded material, an ATP-agarose binding assay was carried out using the purified protein. 24 µg purified TM0288 were incubated for 1h at 4°C with 25 µl ATP-agarose (02065, Sigma), previously equilibrated with 100 µl buffer A. The suspension was applied to a small column, washed twice with 30 µl of buffer A and the protein was eluted by addition of 30 µl buffer A containing 10 mM ATP. The last step was repeated 2 times. In addition, the experiment was carried out in the presence of 5 mM $MgCl_2$. The samples were analyzed by SDS-PAGE (Figure 31). The SDS-PAGE shows that the protein had bound to the ATP-agarose and that the addition of ATP released the transporter of the beads as seen in rows 5-7 and 7-9 of Figure 31. As expected, the addition of Mg^{2+} ions slightly enhanced the binding of the transporter to the beads as concluded by comparing the band intensities of the flowthrough fractions (lane 2 and 9).

Expression and purification of TM0287 was carried out as described for TM0288, except that buffer A contained more salt (500 mM NaCl). Figure 32 shows the purity of the sample after the Ni-column, whereas Figure 33 shows the elution profile during size exclusion chromatography (Superdex 200 HILOAD, 320ml, Pharmacia) and the corresponding fractions, which were analyzed by SDS PAGE. The sample was not as pure as the protein obtained from TM0288 and multiple impurities were visible on the SDS-PAGE.

Further purification with cationic or anionic exchange materials did not improve the protein purity and further analysis was carried out with samples containing these impurities.

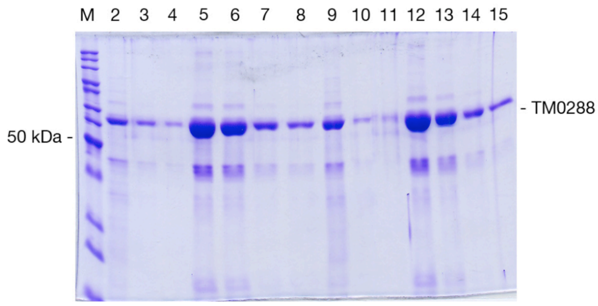

Figure 31: ATP-agarose binding of TM0288. The binding and elution of TM0288 to the ATP-agarose is shown. Lane 1: Marker proteins; lane 2: unbound protein; lane 3-4 wash steps; 5-7: samples after addition of ATP; lane 8: beads after elution. Lanes 9-15: loaded as lanes 2-8 but without the addition of $MgCl_2$.

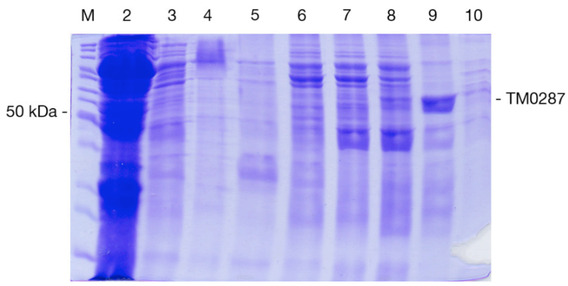

Figure 32: Purification of TM0287. Lane 1: Marker proteins; lane 2: 2 µl of the soluble fraction after cell lysis; lane 3: pellet after solubilization; lane 4: 5 µl of the flowthrough of the Ni-column; lane 5: 250 µl of the wash without imidazole; lane 6: 125 µl of the wash with 10 mM imidazole, lane 7: 25 µl of elution with 50 mM imidazole; lane 8: 25 µl of elution with 100 mM imidazole; lane 9: 25 µl of elution with 250 mM imidazole; lane 10: column material after elution. The small arrow indicates TM0287.

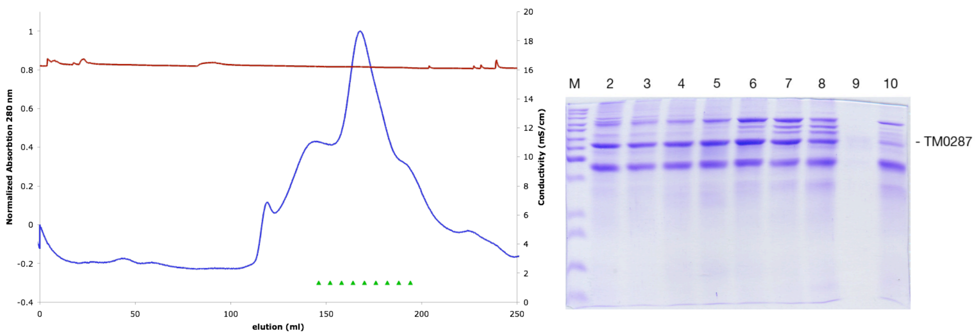

Figure 33: Size exclusion chromatography and SDS-PAGE analysis of TM0287. Size exclusion chromatography (Superdex 200 HILOAD, 320 ml, Pharmacia) indicated higher amounts of aggregates, which eluted in the void volume of the column (~110 ml). Right image: SDS PAGE of the fractions corresponding to those marked with green triangles on the size exclusion chromatography profile. Lane 1: Marker proteins; lanes 2-10: 500 µl of the individual fractions were precipitated and loaded on the gel. Multiple impurities at lower and higher molecular weight were found over the full range of the peak. The arrow indicates TM0287.

The transporter TM0287 showed less stability than TM0288 and precipitated after 2-3 day even at low concentration. Adding 20% (v/v) glycerol or 200 mM sucrose during purification did not improve the stability of the transporter as indicated in Figure 34, which shows the elution of the Ni-column after 2 days incubation at 4°C. The high amount of aggregates that do not enter the gel and the degradation products at lower molecular weight indicate a lower stability compared to TM0288. The addition of sucrose promoted the aggregation and after 2 days of incubation a white precipitate of protein could be observed.

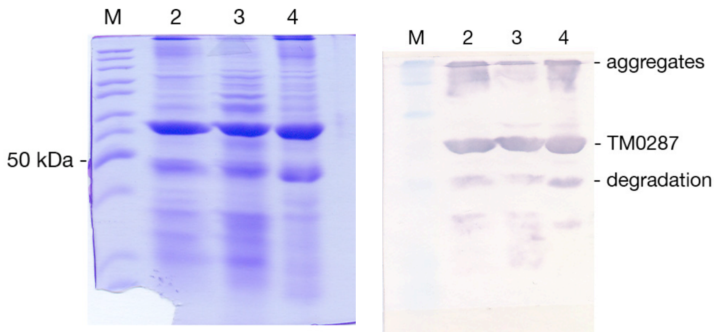

Figure 34: Stability test of TM0287. SDS-PAGE analysis of the fractions of the Ni-column after 2 day storage at 4°C. Lane 1: Marker proteins; lane2: TM0287 purified in 500 mM NaCl; lane 3: TM0287 purified in the presence of 20 % glycerol; lane 4: precipitate of TM0287 purified in presence of 200 mM sucrose. The right image shows the corresponding Anti-His Western blot.

Biochemical Analysis of TM0287 and TM0288

To address the question if the two half transporters TM0288 and TM0287 form a heterodimeric complex, which would be a bacterial analogue of the human TAP1/2 system, or instead form two individual ABC transport systems, a series of experiments were carried out. These experiments examined the full length and truncated constructs of both transporters. The first indication of a possible heterodimer was found upon measuring ATP hydrolysis (see page 169) of the individual transporter and a mixture of both. Individual transporters showed no activity, whereas the activity was increase 10 fold upon mixing of the two transporters (Figure 35). But, since TM0287 could not be purified to homogeneity and the substrate of the transport process is unknown, further experiments to analyze the transport process in more detail could not be carried out.

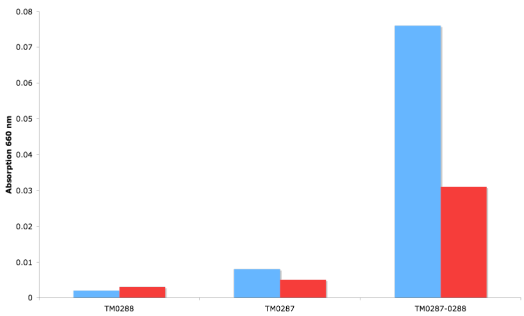

Figure 35: ATPase activity of the transporters TM0287 and TM0288. Both transporters were measured separately and as mixture of both. The mixture showed a higher release of PO_4^{3-} than the individual transporters. The blue bars correspond to twice the amount of protein compared to the measurement that is displayed in red.

However, further investigations on the formation of the heterodimer were carried out. To verify the presence of this complex, the two proteins were separately expressed and purified and subsequently analyzed by size exclusion chromatography either alone or as mixture. Although different detergents and salt concentrations were used, no condition was found in which a clear shift of the elution volume would indicate the presence of the dimeric complex. On the other hand, the formation of the dimeric complex might be prevented, because both the membrane spanning domains were expressed separately and were therefore surrounded by detergent. This question was addressed using constructs without the TMD or the simultaneous expression of both full length proteins in one host.

The NBD of TM0287 (starting at bp 958) and the NBD of TM0288 (starting at bp 1071) were cloned in the vector pat222 which contains an N-terminal 6xHis-tag and a maltose binding protein. An avidin tag was introduced at the C-terminal end of TM0287 and TM0288, resulting in a protein mass of 75 kDa of TM02878 and 70 kDa of TM0288. Proteins were co-expressed in XL1-blue at 18°C and purified by IMAC. The elution of the IMAC was analyzed by size exclusion chromatography (Figure 36). Both proteins eluted at ~ 9.2 ml (Superdex 75, 23 ml, Pharmacia) corresponding to a molecular weight of ~60 kDa. In addition, a 1:1 mixture of both proteins was incubated for 1h and analyzed by size exclusion chromatography. The mixture eluted at the same volume as did the individual proteins, and therefore complex formation of the NBDs of TM0287 and TM0288 could not be observed. The addition of ATP and Mg^{2+}, which should bind in the interface, had no effect upon dimer formation, as shown in Figure 36. The elution profile of TM0288 in presence or absence of ATP of 1 mM ATP and 5 mM $MgCl_2$ was the same.

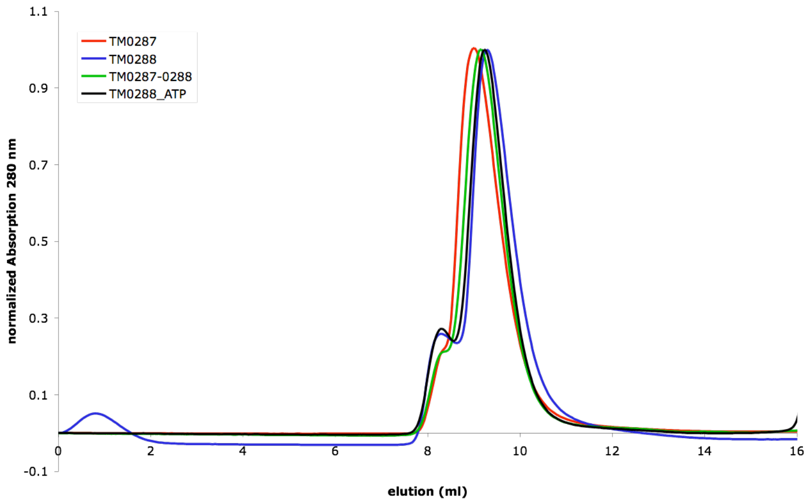

Figure 36: Size exclusion chromatography of the NBDs of TM0287 and TM0288. The picture shows the elution profiles of the NBDs of TM0287, TM0288 and the 1:1 mixture of both proteins in red, blue and green, respectively. Upon mixing of the two individual expressed proteins no experimental evidence was found for the presence of a heterodimer. The elution profile of the NBD of TM0288 in the presence of ATP and Mg^{2+} is shown in black.

The contradicting results on complex formation were further addressed by co-expression of TM0288 and TM0287 in the same host. To ensure that both plasmids were inside the cell, TM0288 was cloned in pET21 vector providing a resistance against ampicillin and TM0287 was used as a pET28-ch construct harboring the gene for the resistance against kanamycin. The construct of TM0288 contained an additional avidin tag to enable detection of both proteins separately. To further ensure that only hetero complexes could be purified, the 6xHisTag at the C-terminal end of TM0288 was removed. Both plasmids were simultaneously transformed into BL21(DE3) *E.coli* cells. The proteins were expressed as described for the transporter TM0288, in the presence of both antibiotics using a controlled pH environment. The fractions from the Ni column were analyzed by Western-Blot and both proteins were visualized by Anti-His or Streptavidin antibodies. Although the presence of TM0288 was verified by crude extract analysis, the inspection of both Western blots indicated that the two transporters did not form a hetero complex. TM0287 was present in the samples of the Ni-column whereas the blot stained with streptavidin showed no band in the corresponding fractions that could be related to TM0288. Overall the results indicated that the two transporters TM0287 and TM0288 do not form a heterodimer, but are rather compounds of two individual ABC transport systems.

C.2.3.2 Proteins Originating from *Escherichia Coli*

ZnuBC

The high affinity zinc uptake system ZnuABC of *E.coli* is used as a second transport mechanism when the environmental concentration of the substrate is reduced. The complete transporter system consists of the periplasmic binding protein ZnuA, the membrane-spanning domain ZnuB and the ATP binding protein ZnuC. Metal loaded ZnuA binds to the TMD of ZnuB in the periplasm of E.coli. The two proteins ZnuB and ZnuC then facilitate the transfer of the zinc ions across the membrane into the cytosol of the bacteria. The three proteins are located in close proximity on the genome of *E.coli*, where ZnuC and B are enconded on the forward DNA strand and ZnuA resides on the reverse strand on the genome. An additional protein Zur, which binds to the bidirectional promoter region of ZnuA and ZnuBC, regulates the expression of the zinc transporter in vivo (Patzer & Hantke, 1998).

Analysis of ZnuB suggested the presence of seven transmembrane helices with the C-terminal end of the protein located in the cytosol (Figure 38) (Daley *et al.*, 2005). Two loops residing in the periplasm are present between helices 2 and 3 and between helix 4 and helix 5. Three longer loops between helices 1 and 2, 3 and 4 or 5 and 6 with a length of 22, 19 and 20 amino acids respectively, are located on the inner side of the membrane.

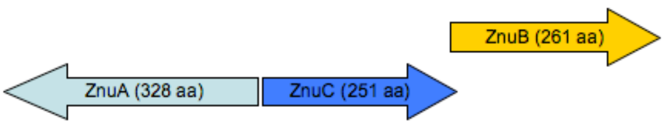

Figure 37: Gene organization of the ZnuABC zinc transport system. The three genes of the zinc uptake system are encoded by base pairs 1939675 to19490607 for ZnuA, 1940686 to 1941441 for ZnuC and 1941438-1942223 for ZnuB. ZnuA is transcribed from the reverse strand whereas ZnuB and ZnuC are located on the forward strand of the DNA. Note that ZnuC and ZnuB have 4 base pairs in common, encoding for the stop codon of ZnuC in frame 1 and the start methionine of ZnuB in frame 2 of the DNA sequence.

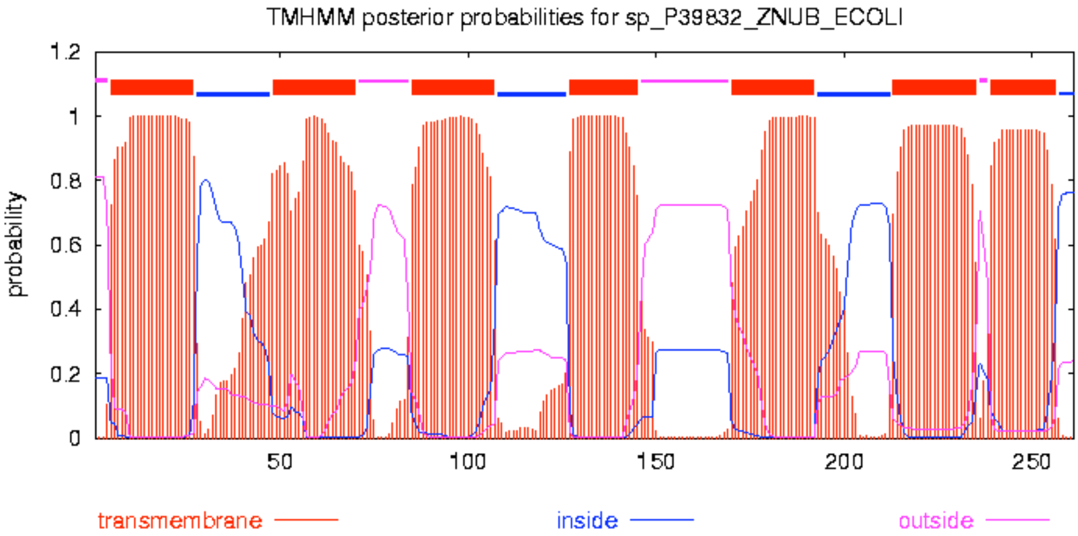

Figure 38: Prediction of the transmembrane helices in the ZnuB. The red bars describe the localization of the transmembrane helices defined by the sequence of ZnuB. The N- and C-termini reside in the periplasm and in the cytosol, respectively. Two loops, located between helices 2 and 3 or helices 4 and 5 are located in the periplasm. Large loops in the cytosol are found between helices 1 and 2, 3 and 4 of between helices 5 and 6. (http://www.cbs.dtu.dk/services/TMHMM/)

The two proteins ZnuB and ZnuC were cloned into the vectors pET21 and pET28 (Table 4) with the 6xHis-tag either at the N- or at the C-terminus of each polypeptide. The two tags generate 4 different combinations depending on which tag position is chosen for each polypeptide. Western blot analysis (Figure 39) of test expressions in *E.coli* BL21(DE3) with all four combinations, showed that the expression of the membrane protein ZnuB was highly affected by the position of the tag. Both proteins were detectable and available in reasonable amounts if the membrane protein was tagged at the C-terminus. An N-terminal tag of the membrane protein resulted in a 10 fold lower expression level. The NBD ZnuB was less affected by the location of the tag and therefore having both proteins tagged at the C-terminal end of the polypeptide resulted in the best expression. Nearly no expression could be detected if both proteins carried their tags at the N-termini of the proteins. Western blot analysis also showed that the expression of the membrane protein ZnuB was much lower than the expression level of the soluble protein ZnuC.

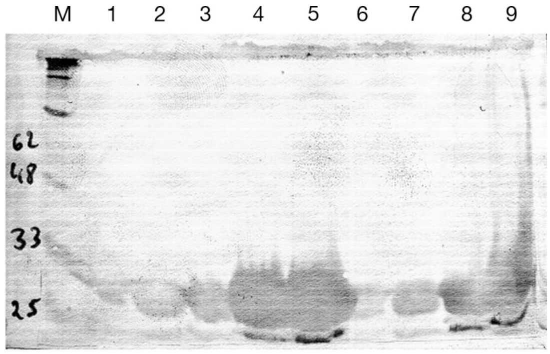

Figure 39: Expression test of the ZnuB-CH and ZnuC-NH. Samples were taken at different time points during growth of the bacteria. To distinguish between both proteins N- and C-terminal were stained separately. The black sharp lines mark the expressed membrane proteins ZnuB whereas the grey dots show the NBD ZnuC. M: marker proteins; lane 1-5: induction with 0.2 mM IPTG, samples are taken after 15`, 30`, 1h, 2h or 3h respectively; lane 6-9: induction with 1 mM IPTG, samples after 15`, 30`, 1h or 2h are shown.

The cells were resuspended (20 mM Tris pH 7.5, 50 mM NaCl, DNAse and protease inhibitors) and opened by FrenchPress. The suspension was centrifuged and the pellet was transferred to a high or low salt solubilization buffer containing 500 mM NaCl or 50 mM NaCl with or without 20% glycerol and 20 mM Tris pH 7.5. To start solubilization, DDM was added to a final concentration of 1% (w/v) and the suspension was gently shaken 1 hour at 4°C. Centrifugation (45000xg, 4°C, 1h) removed non-solubilized material and the supernatant was applied to an IMAC column. The column was washed with 10 column volumes of the buffer, reducing the detergent concentration to 0.03% DDM. A second wash with 50 mM Imidazole was used to remove bound impurities and finally the protein was eluted with 250 mM imidazole. The elutate was concentrated and applied to a size exclusion chromatography column. Analysis of the fractions of the different solubilization procedures revealed that the transporter Znu was not formed correctly. Only one band could be identified, which was related to the ZnuC while the membrane-spanning domain ZnuB was missing in all experiments. Although both proteins could be overexpressed in the cells (Figure 39), the complex could not be purified by the above mentioned protocols. One explanation could be that ZnuB had been inserted incorrectly into the membrane or that the two proteins could not form the correct assembly of the complex when not expressed physically adjacent to each other. Therefore a further construct using the original gene assembly, as it is present in the *E.coli* genome, was generated.

Expression of ZnuC/B as a Single Construct

The two proteins ZnuBC are encoded in the same direction in the genome of *E.coli* sharing 4 base pairs (Figure 37). A single PCR construct (named ZnuC/B) harboring both proteins was generated using the genomic DNA of *E.coli* XL1-blue strain, the 5` primer of ZnuC and the 3` primer of ZnuB. The resulting PCR product was cloned into the modified pET vectors pET28-CH-neu and pET28-NH-neu.

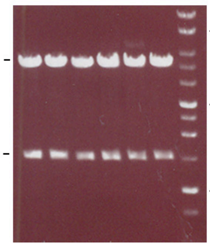

Figure 40: Cloning of ZnuC/B. Double digest of the plasmid DNA obtained after the ligation of the single construct of ZnuC/B with pET28-NH-neu. The plasmids were digested with *BamHI* and *XhoI*. The image shows the backbone at 5.5 kbp and the insert at 1.5 kbp, corresponding to the mass of ZnuC/B. Lane 1-6 : plasmid of the colonies 1-6; lane 7: Marker DNA.

Expression of ZnuC/B was carried out as follows. 500 ml 2YT media were inoculated with cells harboring the plasmid ZnuC/B pET 28-CH-neu. The cells were grown to an $OD_{600} = 0.5$, induced with 0.2 mM IPTG, and protein expression was carried out at 30°C for 4 h. Cells were cracked by FrenchPress in 12.5 ml buffer containing 20 mM Tris pH 7.5, 50 mM NaCl, PMSF, DNAse, RNAse and complete Protease inhibitors (Roche). After centrifugation the pellet was solubilized using 30 ml of 20 mM Tris pH 7.5, 150 mM NaCl and a final DDM concentration of 1%. After incubation at 4°C for 1 h the solution was centrifuged (45000xg, 4°C, 1h) and the remaining supernatant was applied to Chelating Agarose, which was previously loaded with Ni and equilibrated in buffer Z (20 mM Tris, 150 mM NaCl, 0.03%DDM, pH 7.5). The column was washed with 30 ml buffer Z, followed by 20 ml buffer Z containing 50 mM Imidazole. Elution was carried out with 250 mM Imidazole in buffer Z. The eluate was concentrated and analyzed by size exclusion chromatography (Superdex 200, 23 ml, Pharmacia; 20 mM Tris pH 7.5,150 mM NaCl, 0.03% DDM) and SDS-PAGE (Figure 41). The protein had aggregated and eluted after the void volume of the column at ~7.8 ml. But nevertheless, two major bands were visible in the peak and analyzed by tryptic digest and mass spectrometry. The two bands were identified to be ZnuC (lower band) and ZnuB (upper band). In contrast to the experiments with two plasmids, this was the first time that both proteins were found in the same fraction of the elution after size exclusion chromatography. However, the complex was not stable in solution and precipitated after a short period of time. Therefore, the detergent DDM was replaced by Fos-choline-12 or LDAO and the resulting elution profiles and SDS-PAGEs are shown in Figure 42 and Figure 43. With these detergents, the aggregation peak was smaller and two additional peaks at 10.9 ml and 13.0 ml or 11.1 ml and 12.7 ml were found in the presence of Fos-choline-12 or LDAO, respectively. SDS-PAGE analysis of these peaks revealed that only the peak at 11.1 ml contained both proteins whereas the peak at 12.7 ml mainly consists of ZnuB (Figure 43). In contrast, the first peak obtained in the presence of Fos-choline-12 contained neither of the two proteins, but both proteins were found in the peak eluting at 13.0 ml. The proposed complex, formed from two molecules of ZnuB and two molecules of ZnuC would have a molecular weight of ~120 kDa. The elution volume of 13 ml corresponded to an apparent molecular weight of ~50 kDa and therefore, the presence of a correctly formed ABC transport system could not be established. In contrast, in the presence of LDAO, both proteins could be found at an elution volume of 11.2 ml corresponding to a molecular weight of ~110 kDa. This detergent looked, therefore, to be more promising than DDM or Fos-choline-12, however, most of the protein was still found in the fractions equivalent to the void volume of the column. Although numerous variations of buffer compositions and additional detergents were carried out, the aggregation of the proteins could not be sufficiently prevented and therefore the protein could not be subjected to crystallization experiments.

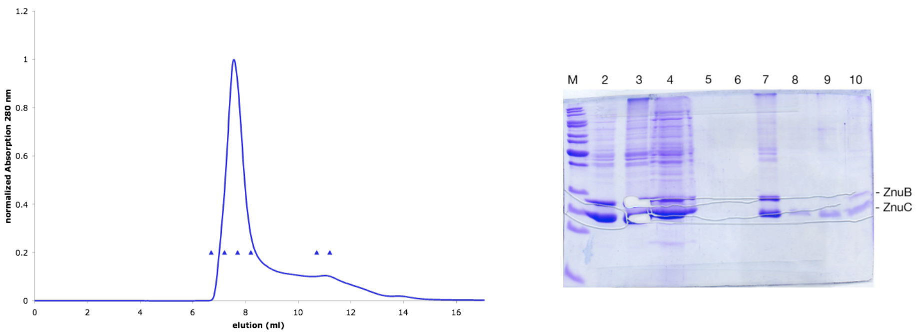

Figure 41: Size exclusion chromatography and SDS-PAGE analysis of ZnuC/B. The elution profile of the size exclusion chromatography column of ZnuC/B in DDM is shown. The fractions of the elution, analyzed by SDS-PAGE are marked with triangles. In DDM the transporter ZnuCB was aggregated and it is found mainly in the void volume of the column. Cut bands were analyzed by tryptic digest and Maldi-Toff to verify the presence of both proteins. ZnuB and ZnuC are indicated. Lane 1: Marker proteins; lane2: elution of Ni-column; 3: precipitate after concentration; 4: supernatant after concentration; lanes 5-10: size exclusion chromatography fractions corresponding to the blue triangles.

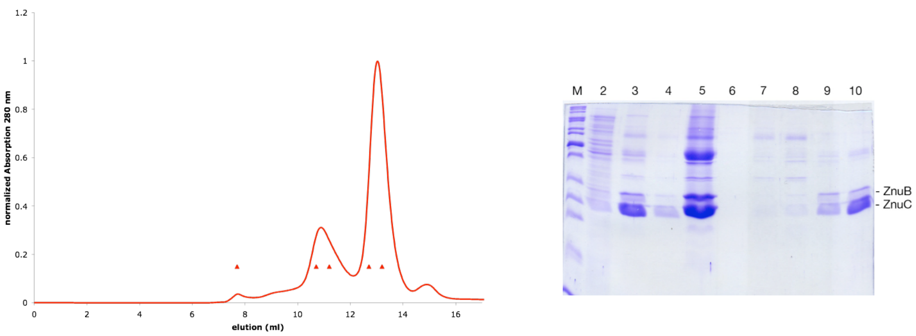

Figure 42: Size exclusion chromatography and SDS-PAGE of ZnuC/B in Fos-choline-12. The elution profile of ZnuC/B in Fos-choline-12 and the purification are shown. SDS-PAGE: lane 1: Marker proteins; lane 2: wash with 50 mM imidazole; lane 3-4: elution with 250 mM imidazole; lane 5: concentrated sample before size exclusion chromatography; lane 6-10: size exclusion chromatography fractions corresponding to the red triangles.

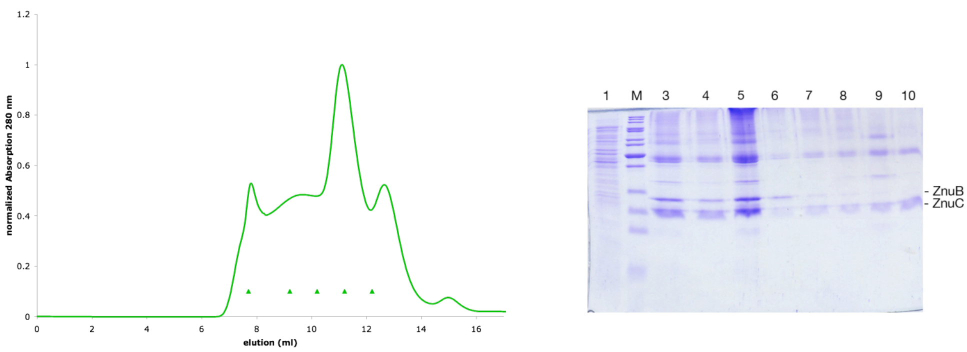

Figure 43: Size exclusion chromatography and SDS-PAGE of ZnuC/B in LDAO. The elution profile of ZnuC/B in LDAO and the purification are shown. SDS-PAGE: lane 1: wash with 50 mM imidazole; lane 2: Marker proteins; lane 3-4: elution with 250 mM imidazole; lane 5: concentrated sample before size exclusion chromatography; lane 6-10: size exclusion chromatography fractions corresponding to the green triangles.

YDDA

The protein YDDA from *E.coli* consists of 561 amino acids with a molecular weight of 64985 Da. This protein, as yet, belongs to the class of ABC transporters with unknown function. The membrane spanning domain and ATP binding domain are fused together, and YDDA, therefore, belongs to the so called half-transporters. Analysis of the topology of YDDA reveals five transmembrane helices with the C-terminal end of the protein located in the periplasm (Figure 44). Since the ATP binding C-terminal region of the polypeptide must reside in the cytosol to maintain its function, this prediction is false. Daley et al (Daley *et al.*, 2005) fixed the C-terminus to the cytosol and the prediction changed to 4 transmembrane helices.

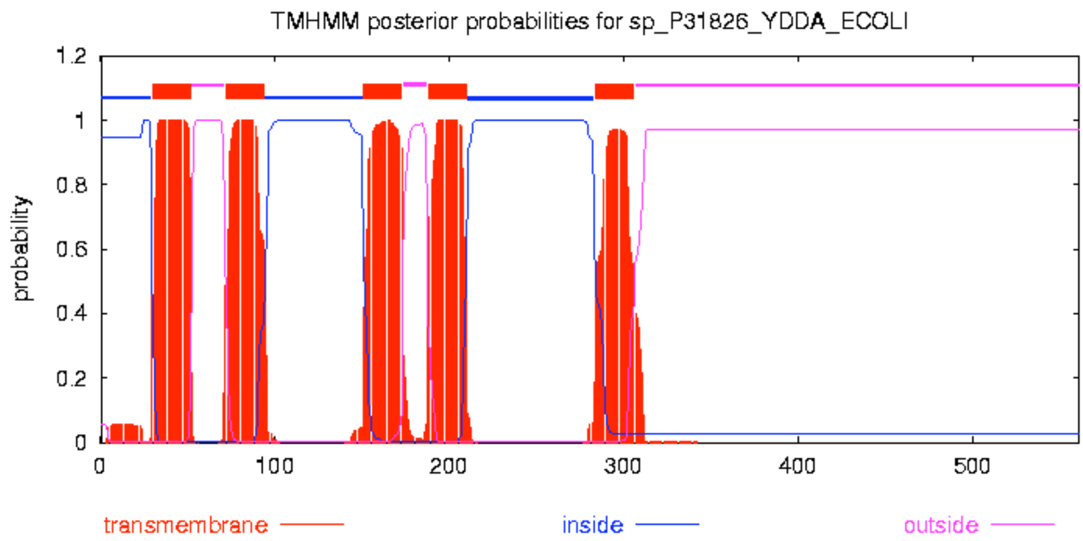

Figure 44: False predicted transmembrane helices of YDDA. The prediction results in 5 transmembrane helices with the C-terminus of the protein located in the periplasm of the cell. Since the C-terminus must reside inside of cell to maintain its function as an NBD, this prediction is false. Fixing the C-terminus to the cytosol, Daley et al (Daley *et al.*, 2005) reported four membrane-spanning helices for the protein YDDA.

Amplifying the gene of YDDA using genomic DNA of *E.coli* XL1-blue as the template, generated a PCR construct, which was cloned into the two vectors pET28-CH and pET28-NH using *BamHI* and *XhoI* as restriction enzymes. Bl21(DE3) cells were grown to an OD_{600} = 0.7 at 37° in 2YT media and protein production was started with 0.2 mM IPTG and carried out at 18°C over night. The cells were harvested, resuspended in lysis buffer (20 mM Tris pH 7.5, 150 mm NaCl, 1 ug/ml DNAse, 1 mM PMSF and protease inhibitors containing EDTA (Roche)) and opened by FrenchPress.

The solution was centrifuged (1 h, 40000xg, 4°C) and the remaining pellet was solubilized for 1.5 h at 4°C in lysis buffer containing 1% (w/v) DDM. To avoid interference with the metal chelate column, EDTA was not added to the solubilization buffer. Insoluble material was removed by centrifugation (1 h, 4°, 100000 x g) and the supernatant was loaded on 1 ml Protino Ni-IDA (Machery-Nagel) beads. The column was washed with 40 column volumes (CV) of buffer YA (20 mM Tris pH 7.5, 150 mM NaCl, 0.03% DDM), 20 CV of buffer YB (20 mM Tris pH 7.5, 300 mM NaCl, 0.03% DDM) and 20 CV of buffer YC (20 mM Tris pH 7.5, 150 mM NaCl, 0.03% DDM, 2.5 mM Imidazole). Protein was eluted with 1 x 0.5 ml and 4 x 2 ml of buffer YE (20 mM Tris pH 7.5, 150 mM NaCl, 0.03% DDM, 150 mM Imidazole). SDS-PAGE analysis of the purification process is shown in Figure 45. Fractions 2-5 of the elution were concentrated and applied to a size exclusion chromatography column (Superdex200, 23 ml, Pharmacia) equilibrated with 20 mM Tris pH 7.5, 150 mM NaCl, 0.03% DDM. The main fraction of the protein eluted after 10.5 ml corresponding to a molecular weight of 200 kDa (Figure 46).

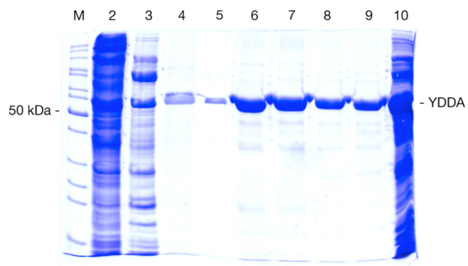

Figure 45: IMAC purification of YDDA. The gel shows the enrichment of the protein during purification. Lane 1: marker proteins, lanes 2-4: 1 ml of the corresponding wash steps are loaded; lanes 5-9: 10 µl of the elution 1-5, lane 10: resin.

Fractions containing YDDA were concentrated and crystallization trial set up in the 96-well plate format. Although the protein concentration varied from 5-20 mg/ml, the major observation was a heavy precipitation of the protein immediately after the initial set up of the crystallization experiment, and therefore further detergents were tested to improve the stability of the protein. Two strategies were applied during these experiments.

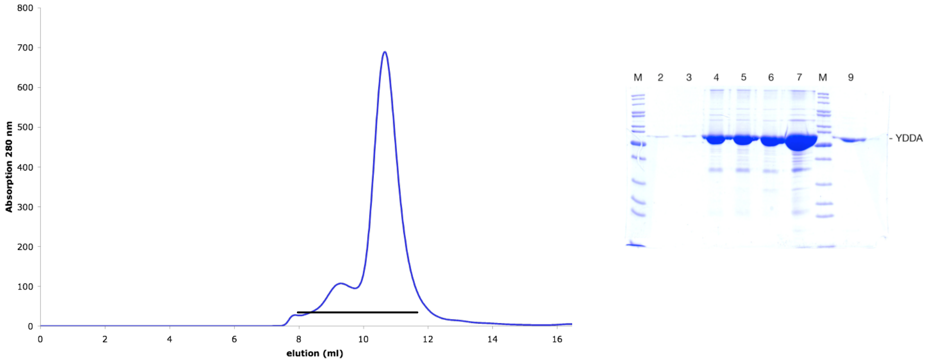

Figure 46: Size exclusion chromatography and SDS-PAGE analysis of YDDA. Elution fractions from the Ni-column were concentrated and applied to a size exclusion chromatography column. The main protein peak eluted at 10.5 ml corresponding to an apparent molecular weight of 200 kDa. The fractions corresponding to the two peaks were loaded on a SDS-PAGE. Lane 1 and 8: Marker; lane 2-7 and 9 correspond to elution volume 8.7 to 12.3 ml in steps of 0.5 ml (black line).

Detergents with high CMC values were not able to solubilize the transporter from the membrane, and, therefore, these detergents were introduced at the stage of the IMAC column whilst the protein was initially solubilized with DDM. In contrast, detergents with low CMC values were able to extract the protein from the membrane and therefore detergents like Fos-choline 14 or Fos-choline 16 were used during the initial steps of the purification. Size exclusion chromatography analysis of YDDA in the different detergents is shown in Figure 47. In the presence of DDM and Fos-14, the protein eluted as a sharp peak whereas the elution profiles in Fos-12 and DM showed a broader peak. In contrast, the detergents Fos-16 and Fos-9-iso destabilized YDDA and higher amounts of aggregates were observed.

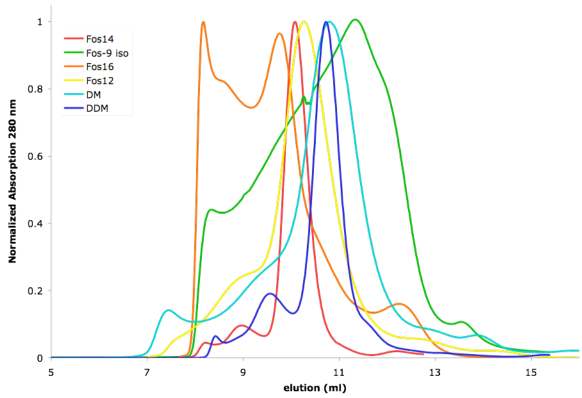

Figure 47: Behavior of YDDA in different detergents during size exclusion chromatography analysis. See text for details.

Biochemical Characterization of YDDA

Reconstitution of the YDDA into liposomes followed by an activity assay of the transport process could not be exploited since the substrate of YDDA is still unknown. Therefore, activity of the protein was measured upon ATP binding. 10 µg protein, which had been solubilized in DDM, were incubated for one hour at 4° with 5 mg ATP-agarose (Sigma Aldrich). The beads were transferred to a small column and washed twice with 20 µl buffer AB (20 mM Tris, 150 mM NaCl, 0.03% DDM, pH 7.5). Elution of the protein was carried out with 3 x 20 µl buffer AB containing 10 mM ATP. Samples were analyzed by SDS-PAGE (Figure 48). As shown in Figure 48, the protein did not bind to the ATP-agarose and all protein was found in the flowthrough of the column. Although the experiment was repeated several times, including the solubilization of YDDA in the mild detergent digitonin, variations of the ATP anchor point to the agarose and the spacer length, no ATP binding of the protein could be observed.

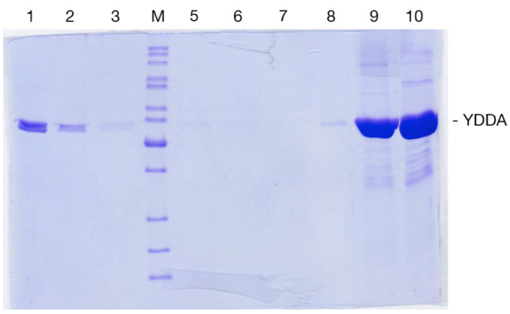

Figure 48: ATP-binding assay of YDDA. The SDS-PAGE analysis of the binding experiment with ATP agarose (attached at C8, 9 atoms spacer) is shown. Lane 1: flowthrough; lane 2-3 wash 1+2; lane 4: marker proteins; lanes 5-7: elution 1-3; lane 8: resin after elution, lane 9: YDDA positive control.

Since correct folding of the protein by binding to ATP could not be established it was tried to verify the presence of folded material in the detergent environment by an FT-IR experiment. YDDA was purified in the presence of DDM and concentrated to 22 mg/ml. 20 µl of the sample were applied and an attenuated total reflectance (ATR) experiment was carried out (Tensor 27 FT-IR; Bruker). The spectrum obtained in the region of the amide I band is shown in Figure 49. Comparison of the spectrum with a set of spectra obtained from proteins with known 3D structures resulted in a prediction of 47% α-helices and 11% β-sheets, which is in excellent agreement with the prediction of the ABC transporter BtuCD that has 45% α-helices and 17% β-sheets.

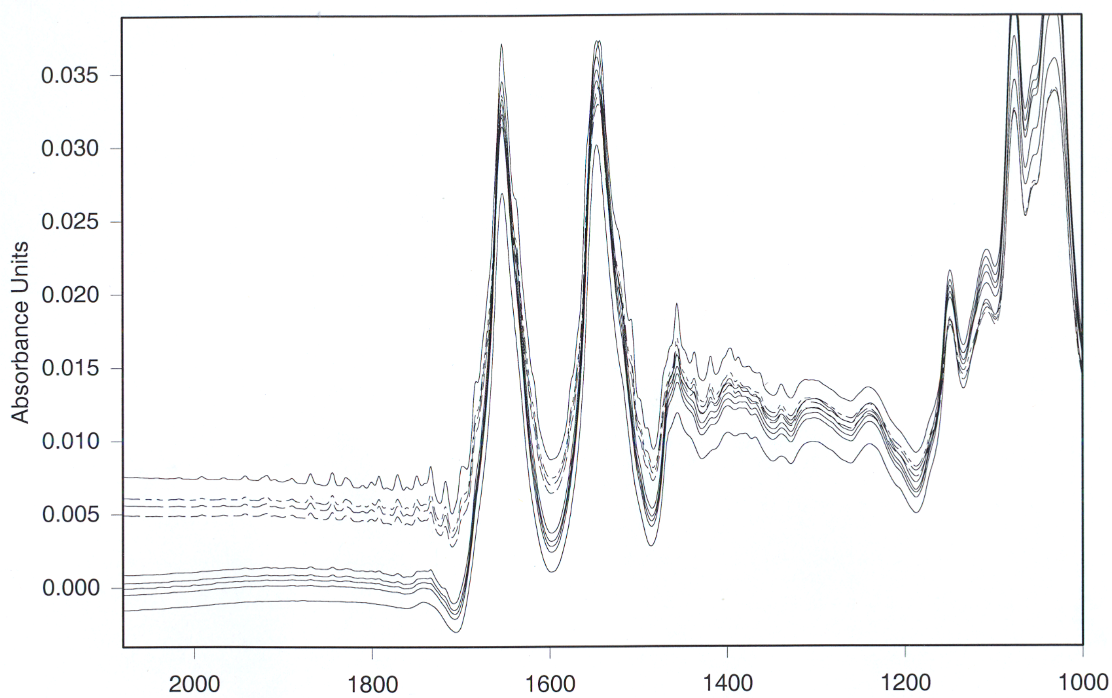

Figure 49: FT-IR spectra of YDDA. The image shows the amide I region (1700 - 1600 cm-1) that is mainly influenced by the C=O stretching vibration of the peptide bond. Comparison with a set of known structures is shown and revealed 47% of α-helices and 11% β-sheet content of YDDA.

To determine the absolute mass and the oligomeric state of the transporter in solution, a static light scattering experiment was carried out (miniDAWN, Wyatt). The protein was purified according to the general purification protocol and applied to a size exclusion chromatography column (Superose 6, 23 ml, Pharmacia) coupled to the miniDAWN (triangular light scattering detector, Wyatt) and OPTILAB (interferomertric Refractometer, Wyatt) instruments. An internal calibration run with 100 µl of 2.5 mg/ml BSA was used to detect the correct settings of the dead volume and delay of the system. The major peak of the elution profile contained protein with a polydispersity (M_w/M_n) of 1.000 ± 0.005 and a molecular weight of 250 kDa which is in good agreement to a homodimeric protein (2x 65 kDa) plus the attached micelle of DDM with a mass of 70 -100 kDa. In summary, the biochemical investigations on YDDA showed that the protein is folded and forms a homodimeric complex.

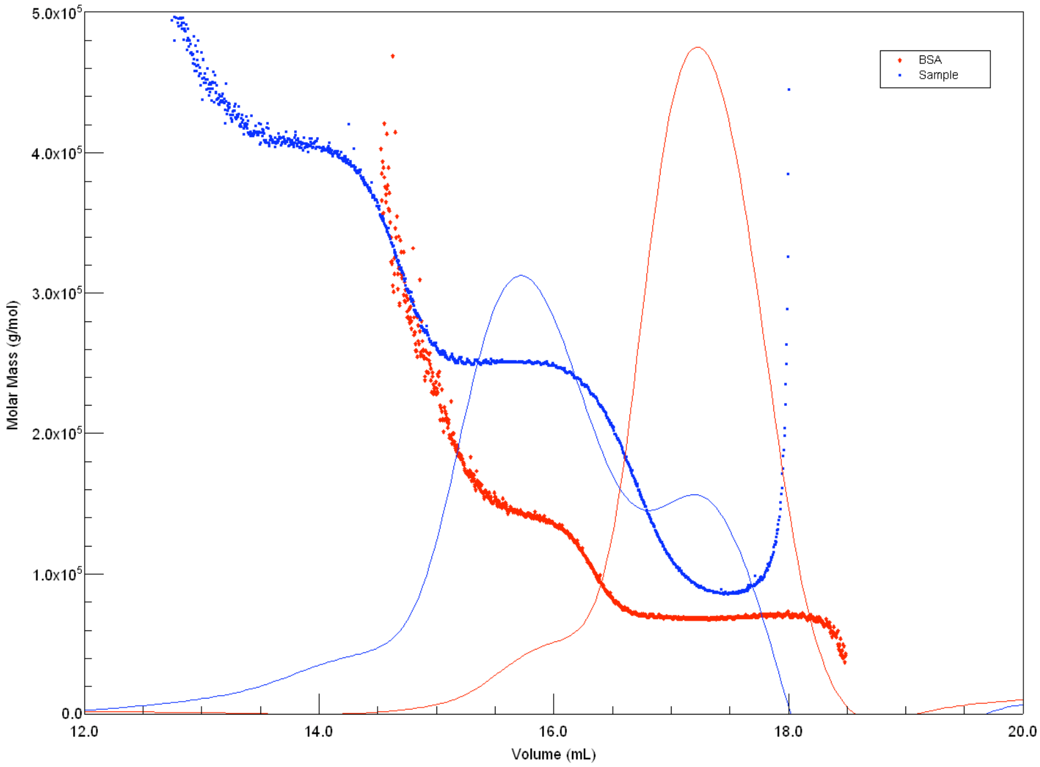

Figure 50: Static light scattering of YDDA. The UV_{280} signal (solid line) and the distribution of the molar mass (dotted line) vs. the elution volume are shown of YDDA and BSA in blue and red, respectively. The main peak of YDDA consists of a monodispers 250kDa species.

AraG

The L-arabinose transport system of *E.coli* consists of three proteins named AraF, AraG and AraH. The periplasmatic protein AraF, which binds the solute in the periplasm, targets the sugar to the transporter. The membrane spanning domain AraH, which forms 10 helices across the membrane, builds up the pore where the sugar crosses the lipid bilayer. The ATP binding protein AraG provides the energy used by the transport process. This protein is special among the selected ABC transporters, since it contains a fusion of two NBDs. The special arrangement of the two domains on one polypeptide might reveal answers to the question of how the two nucleotide binding domains are arranged in an active transporter. Therefore, the purification and crystallization of the NBD AraG of the L-arabinose transport system was pursued.

Expression and Purification of AraG

Freshly transformed *E.coli* BL21(DE3) cells, harboring the full length cDNA of AraG in a pET21-CH vector, were used to inoculate 2 ml LB media. The pre-culture was transferred after 4h to a second pre-culture with a volume of 25 ml LB media. This second pre-culture was used to inoculate 2 x 500 ml LB-media containing the antibiotic ampicillin. Cells were grown to an OD_{600}=0.7 at 37° and protein expression was induced by the addition of 0.2 mM IPTG and the cells were left at 18°C o/n. Cells were harvested by centrifugation and stored at -20°C.

The cells were suspended in 25 ml of buffer PA (50 mM Tris, 50 mM NaCl, pH 8.0) containing 1 mM PMSF and DNAse and finally opened by FrechPress. Cell debris was removed by centrifugation (40000 x g, 4°C, 1h) and the supernatant was applied to a 4 ml Ni-NTA IMAC column. The column was washed with 50 ml PA followed by 30 ml PA containing 50 mM imidazole. The protein was eluted with 7 mL PA containing 250 mM imidazole. The elution was applied to a size exclusion chromatography column (Superdex 75 26/60, 320 ml, Pharmacia) and isocratic elution was carried out in buffer PA (Figure 51).

The main fraction of the protein eluted at ~150 ml corresponding to a molecular weight of 50 kDa, which is in good agreement with the theoretical mass of 56 kDa. SDS-PAGE analysis showed that the two purification steps produced protein still containing minor impurities.

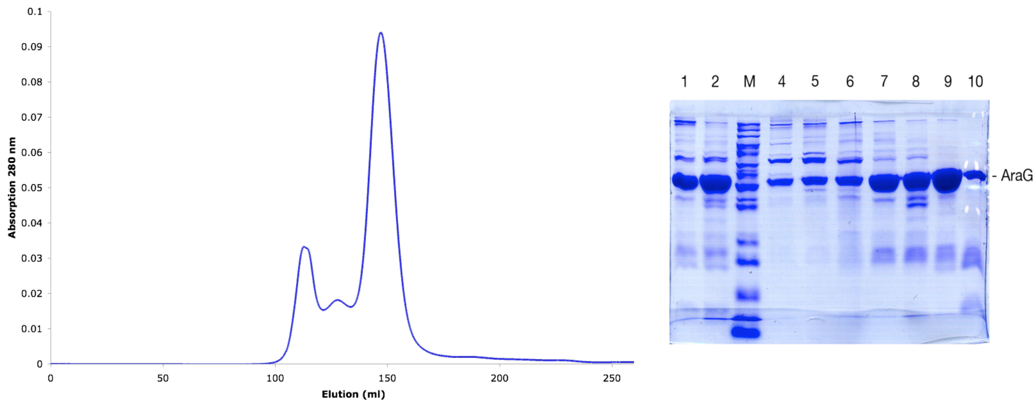

Figure 51: Size exclusion chromatography and SDS-PAGE analysis of AraG. The main peak corresponds to a molecular weight of 50 kDa, which is in good agreement with the theoretical mass of 56 kDa. Lanes 1-2: 2 µl of the elution of IMAC column (not boiled and boiled); lane 3: marker proteins; lanes 4-5: elution 1 and 2 corresponding to the first peak showing aggreagated protein; lane 6: elution 3 corresponding to the middle peak; lanes 7-10: elution 4, 4 (boiled), 5 and 6 corresponding to the main peak in the size exclusion chromatography. 30 µl of the individual samples were loaded.

Therefore, the protein was further purified using a MonoQ column (8 ml, Pharmacia). Fractions of the elution corresponding to the main peak of the size exclusion chromatography were diluted 1:1 with 50 mM NaCl and applied to the column. Protein was eluted using a 50 to 150 mM NaCl gradient in 10 CV.

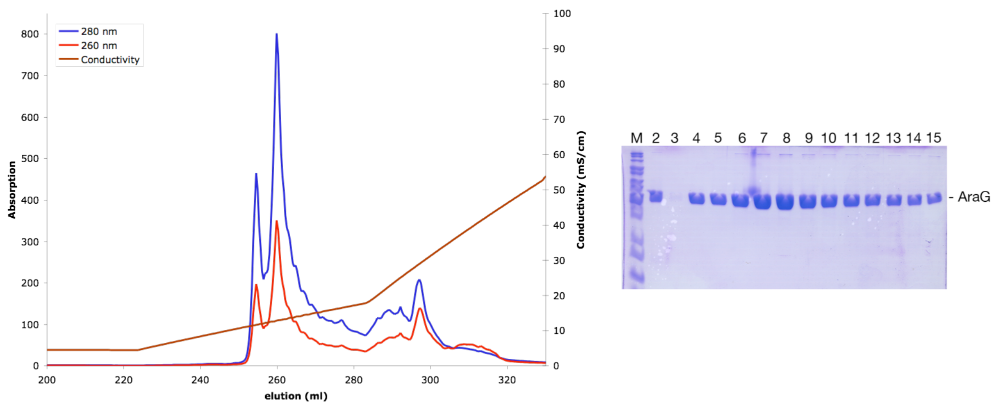

Figure 52: IEX and corresponding SDS-PAGE analysis of AraG. The chromatogram of the MonoQ run is shown. AraG elutes in two peaks that contain pure protein. Lane 1: marker proteins, lane 2 combined fractions of the size exclusion chromatography run, lanes 3-15: elution corresponding to the two main peaks.

SDS-PAGE analysis of the samples showed that the protein was pure and no other contaminations were present in the samples (Figure 52). Interestingly, the elution profile of AraG showed two peaks, which both contained pure protein; analytical IEX chromatography in the presence of 5 mM ATP and 5 mM $MgCl_2$ revealed a difference in their ability to bind the supplemented nucleotide. The absorption at 260 nm is dominant within the second peak, whereas the first peak showed no additional signal at 260 nm, indicating that the protein that eluted in peak 2 was able to bind the nucleotide and ATP, whilst the protein in peak 1 was not. It was therefore, concluded, that peak 1 either contained misfolded protein or protein in a conformation unable to bind ATP (Figure 53). The protein corresponding to the active fractions of AraG was concentrated up to 5 mg/ml and crystallization trials were set up. The crystallization process was complicated by the fact that the protein AraG was prone to aggregation. Reducing the protein concentration or adding inhibitors, e.g. AMP-PNP, ADP-VO_4^{3-}, could not resolve the problem and no suitable initial crystallization conditions have been found so far.

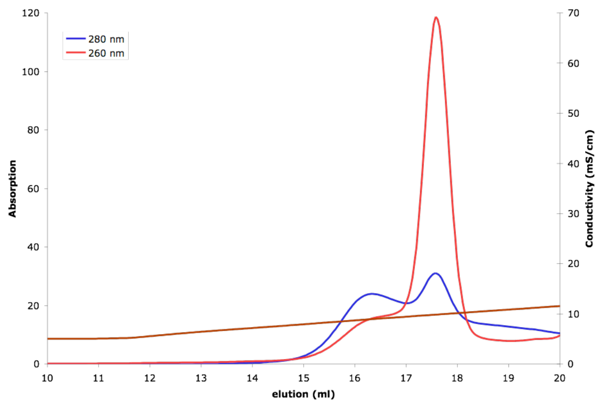

Figure 53: Binding of ATP to AraG. Analytical IEX in the presence of ATP and $MgCl_2$ revealed that the protein under peak 2 was able to bind ATP whereas peak 1 contains an inactive form of AraG.

C.2.3.3 Crystallization of ABC transporters

The purified protein samples of TM0288, TM0287 or YDDA were concentrated to perform crystallization experiments. Various screens including wide ranges of polyethylene glycol, buffer and salt conditions were performed. ATP, ADP, ADP and vanadate (VO_4^{3-}) or AMP-PNP were used to enhance the homogeneity of the sample and enforce the dimerization of the NBD. No suitable conditions for crystal growth of one of the full transporters have been found so far for the selected proteins.

C.3 Discussion and Conclusion

The results described in this part of the thesis demonstrate the successful selection, purification, and partial characterization of ABC transporters originating from different species.

The initial work on ABC transporters included the work on mBsep, a mammalian bile salt export pump. Although, the protein could be functionally expressed in insect cells, the total yield of expressed transporter was low. Expression tests, in combination with silver stained SDS-PAGE and Western blot analysis using an Anti-his antibody, showed no differences in the band intensities. Structural and biochemical characterization of a protein demands milligrams of pure and homogenous sample, and therefore the solubilization and purification of the protein was not carried out since the quantities of protein required for the experiments could not be produced. Therefore, we focused our work on prokaryotic ABC transporters.

The large number of different ABC transporter sequences available in the databases complicated the selection process. Thus, additional criteria were applied to enable a selection of a subset of transporters, with the focus on half transporters and transporters where the function is known. Cloning of the transporters was straightforward but although various PCR protocols were employed, it was not possible to amplify the DNA of the genome of *B. subtilis*. The problem, therefore, is most likely the high GC content of the DNA in these bacteria. However, all other selected transporters were cloned and expression tests were carried out. The expression test showed that, in general, the pET20 constructs were not very successful and the cells often died at early stages during the expression tests. In contrast, using pET21 and pET28 constructs, all cloned transporters produced viable cells. The difference between the two systems pET20 and pET21/pET28 is a different in the regulation of the transcription of the plasmid. Transcription in the vectors pET21/pET28 is better regulated than in pET20 and, therefore, the intrinsic expression of the ABC transporter is reduced, which leads to lower mortality of the *E.coli* cells. Overall, the expression tests revealed that there is no expression strain that is ideal for all ABC transporters and that optimal expression for each protein needed slightly different culture conditions. In contrast, there was a preference for the position of the His-tags. Expression of half transporters was increased, or even only detectable, if the His-tag was placed at the C-terminus of the protein. Although, the individual TMDs of *E.coli* were less affected by the position of the His-tag, again these proteins showed a higher expression yield with a C-terminal His-tag. A single exception was found, the TMD of the Xylose transporter of *E.coli* could only be expressed with an N-terminal His-tag. Membrane proteins are synthesized by membrane associated ribosomes that interact with the translocon machinery of the cell (Hessa *et al.*, 2005, Mackinnon, 2005, Mitra *et al.*, 2005, Van den Berg *et al.*, 2004). The properties of N-terminal nascent chain of the protein determines further processes and thus the additional N-terminal His-tag might interfere with the correct insertion of the transporter into the cell membrane and might explain why the position of the His-tag at the C-terminus of the protein is favored.

Solubilization and purification could be achieved for most transporters using one of the following detergents: DDM, LDAO or Fos-cholines. None of the expressed transporters were resistant to solubilization, however the stability of the proteins in the different detergents varied and therefore had to be explored for each transporter individually. Shorter, and therefore harsher, detergents were prone to induce aggregation of the proteins. Although all transporters could be solubilized and partially purified with IMAC, most of them could not be recovered after size exclusion chromatography analysis. Exceptions were the transporters of *T. maritima* TM0288 and TM0287, the *E. coli* transporters YDDA, YOJI and ZnuC/B. In this respect, giving preferences to half-transporters was worthwhile since 4 out of 5 of the purified transporters belong to this class. On the other hand, it is hard to judge if a protein, which elutes at 130 kDa of a size exclusion chromatography column e.g. TM0288, has formed the correct dimer assembly, or if the protein is eluting as a non-physiological monomer.

The formation of a hetero dimer of the two proteins of *T. maritima* TM0288 and TM0287 is still ambiguous. As it was shown, a mixture of both proteins hydrolyzed ATP at higher rates than each individual protein. However, the presence of the heterodimer could not be established by other biochemical investigations and this issue could not be resolved. Furthermore, hydrolysis of the nucleotides is not correlated with transport process. Low rates of intrinsic ATPase activity have been reported for solubilized NBDs of ABC transporters (Benabdelhak *et al.*, 2005, Greller *et al.*, 2001). Once the nature of the transported substrate is known, the question of the dimer assemblies of TM0287 and TM0288 could be solved, since transport experiments with reconstituted proteins in liposomes and the appropriate substrates could be carried out.

In contrast to TM0288 and TM0287, the protein YDDA of *E.coli* eluted at the approximate calculated elution volume and the presence of the dimer was verified by static light scattering. However, although the presence of folded material was established by FT-IR, the protein was unable to bind the substrate ATP, which is in contrast to TM0288 that did bind to ATP-agarose. In addition, it was not possible to determine the intrinsic ATPase activity of the protein in detergent micelles. A complete or partial inhibition has been described for the multi drug resistance protein PGP (Callaghan *et al.*, 1997, Orlowski *et al.*, 1998). Here it was shown, that the detergent, which was present in the experiment below the CMC concentration, had not only an inhibitory effect on ATPase activity but also reduced the binding of the substrate vinblastine. Interestingly, the effect could be partially reversed by the addition of lipids to the protein, indicating that lipids are essential for the function of this ABC transporter. Although no functional assay of YDDA could be performed, the protein was subjected to crystallization trials.

The purification of the transporter ZnuC/B showed the importance of the construct design. Although both proteins were overexpressed from two plasmids, the transporter did not form correctly in the expression host *E. coli*. The situation changed when both proteins were incorporated on one plasmid.

The full transporter could be partly purified with IMAC but failed subsequently in size exclusion chromatography experiments. Nevertheless, only one of the two proteins carried the His-tag and therefore it can be concluded that a correct assembly of the two proteins was at least present during the first purification steps. This information is important for future work on this particular transport system.

In summary, 28 ABC transporters were selected for our work, resulting in 120 constructs that were analyzed for expression. 15 transporters (54%) could be expressed in sufficient amounts. Of these, we were able to solubilize 5 transporters (18%) in different detergents. 2 transporters (7%) TM0288 and YDDA showed the expected molecular weight during size exclusion chromatography analysis, and were therefore, subjected to crystallization trials. However, up to now, no initial crystallization conditions for these 2 transporters could be determined.

The low abundance of eukaryotic ABC transporter in natural sources, the problems during production of these proteins, (Linton & Higgins, 2002) led us to the here presented strategy of using prokaryotic homologues. However, although two prokaryotic ABC transporters could be purified and obtained in sufficient amounts, the high number of sequences that failed during screening raised concerns and several improvements might be advantageous:

(i) Selection. Although our experience showed that it was more straightforward to use half transporters, most of the published high resolution structures are derived from transporter consisting of multiple proteins (Hollenstein *et al.*, 2007, Locher *et al.*, 2002, Pinkett *et al.*, 2007). Exceptions are Sav1866 (Dawson & Locher, 2006) and the low resolution structures of Pgp and ABCG2 (McDevitt *et al.*, 2006, Rosenberg *et al.*, 2005) obtained by cryo electron microscopy. Since bacterial genomes contain many transporters that are made of multiple proteins, it is obvious that the restriction to half transporter leads to a limit variability of sequences available for analysis.

(ii) Construct design. As seen for the zinc transporter ZnuC/B, the expression of the transporter originating from one plasmid is advantageous. Therefore, all transporters that consist of two or more proteins should be cloned as a single construct.

(iii) New sequences. Although the *E.coli* proteins are annotated and their functions mostly known, their expression and purification yields were not superior to other proteins originating from other bacteria. Therefore, there should be no restriction of the origin of an ABC transporter during selection process and all available genomes should be analyzed for the presence of these transporters.

(iv) Mutations. Studies with single site mutations showed effects on binding of the nucleotide, dimerization of the NBDs and transport processes (Benabdelhak *et al.*, 2005, Chen *et al.*, 2003, Davidson & Sharma, 1997, Smith *et al.*, 2002, Yuan *et al.*, 2001). The stabilization of the NBD dimer is of special interest, since it stabilizes the protein in presence of ATP (Smith *et al.*, 2002). The tremendous effects of these mutations will be exploited in the near future for the NBD of the L-arabinose transporter from *E.coli*, which shares two NBDs on the single gene *araG*. This structure would represent a close homologue of the mammalian ABC transporters.

Chapter D - CitS

D.1 Introduction to the Sodium Citrate Symporter CitS

In contrast to the ABC transporters, which utilize ATP to create an active transport process across the membrane, the family of the 2-hydroxycarboxylate transporter (2HCT) uses the electrochemical salt or proton gradients to transport the solutes across the membrane.

The secondary transporters belonging to the 2HCT family are found exclusively in bacteria and transport substrates such as citrate, malate, and lactate (Sobczak & Lolkema, 2005c), which are all 2-hydroxy derivatives of tri-, di-, and monocarboxylates (Figure 54).

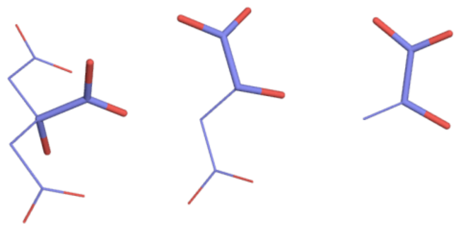

Figure 54: Substrates of the 2HCT transporter family. The tri-, di-, and mono-carboxylates citrate, malate and lactate are show from the left to the right, respectively. The 2-hydroxy carboxylate group, providing the name of this transporter family is shown in bold.

The family members are further subdivided according to their transport properties into exchanger and symporter. Well-studied symporters are the Na^+/citrate symporter CitS of *K. pneumoniae* or Na^+/malate MaeN symporter of *B. subtilis*. Known exchangers are CitP of *Leuconostoc mesenteroides* or MleP of *Lactococcus lactis*, which transport citrate or malate into the cytosol and thereby export lactate. The substrates citrate and malate are precursors in the citrolactic and malolactic fermentation pathways where they are degraded to lactate. The exchange of divalent citrate or malate to monovalent lactate generates a membrane potential that is used by the bacteria as secondary metabolic energy.

D.1.1 Anaerobic Citrate Metabolism: Citrate Fermentation

Several species of enterobacteria are capable of using citrate as a carbon and energy source, whereas most *E.coli* strains are not able to utilize this carbohydrate. Under aerobic conditions *E.coli* lacks an appropriate transport system to facilitate the uptake of the solute citrate into the cell. *E.coli* is however, able to ferment citrate under anaerobic conditions, but the fermentation of citrate to succinate requires an oxidizable co-substrate such as glucose or glycerol and the citrate succinate antiporter CitT (Pos *et al.*, 1998). In contrast, the closely related enterobacterium *Klebsiella pneumoniae*, one of the best-studied organisms with respect to citric fermentation (Bott *et al.*, 1995), is able to grow on citrate under both aerobic and anaerobic conditions.

Three different transport proteins involved in citrate transport have been identified in *K. pneumoniae*. The citrate/proton symporter CitH is expressed during aerobic growth and transports citrate (H-citrate^{2-}) in co-transport with three protons (van der Rest *et al.*, 1992). The CitS and CitW transporters are both expressed under anoxic conditions and play a major role in the citrate fermentation of *K. pneumoniae*. CitW was recently identified as a citrate/acetate antiporter (Kästner *et al.*, 2002). The protein CitS, on the other hand, is the transporter that facilitates citrate uptake coupled to the transport of sodium ions. Figure 55 shows a summary of the citrate fermentation of *K. pneumoniae* during anaerobic growth. Imported citrate is degraded to oxaloacetate and acetate by the enzyme citrate lyase. This 550 kDa complex consists of six copies of each of the α, β and γ subunit (Schneider *et al.*, 2000). The membrane bound enzyme oxaloacetate decarboxylase sodium pump (OAD) further converts the resulting product oxaloacetate. This enzyme is a multi subunit protein consisting of three subunits α, β and γ in a 1:1:1 stoichiometry. The α subunit catalyzes the carboxyl transfer from the substrate oxaloacetate to generate pyruvate and thereby modifies the prosthetic biotin group to carboxybiotin. Decarboxylation of carboxybiotin is facilitated by the β-subunit of OAD in a sodium dependend reaction and two sodium ions are consequently translocated from the cytosol into the periplasm (Dimroth & Thomer, 1990). The enzyme pyruvate formate lyase (PFL) catalyzes the conversion of pyruvate and CoA to acetyl-CoA and formate. Interestingly this homo-dimeric protein uses a glycine radical in the enzymatic reaction (Becker *et al.*, 1999). Formate is decarboxylated by the enzyme formate hydrogen lyase to produce carbon dioxide and hydrogen. The hydrogen is used by a membrane bound hydrogenase to convert $NAD(P)^+$ to NAD(P)H and therefore provides reducing equivalents to the metabolism (Steuber *et al.*, 1999, Kästner *et al.*, 2002).

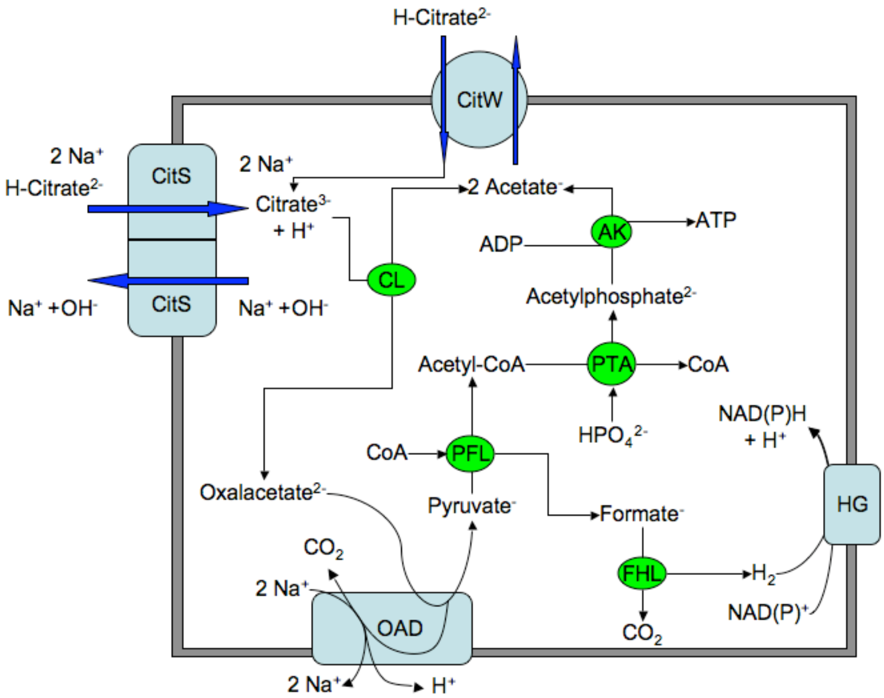

Figure 55: Citrate Fermentation in *Klebsiella pneumoniae*. Citrate is imported into the cell by the protein CitS in a sodium dependent transport process. The 2-hydroxycarboxylate is degraded by citrate lyase (CL) to acetate and oxalacetate that is further decarboxylated to formate and pyruvate by the oxalacetate decarboxylase Na^+ pump (OAD). The enzyme pyruvate formate lyase (PFL) metabolizes pyruvate to formate and Acetyl-CoA. Formate is decarboxylated and the resulting hydrogen is used by a membrane-bound hydrogenase (HG) to reduce $NAD(P)^+$. The acetyl group is transferred to phosphate by the enzyme phosphotransacetylase (PTA). The resulting product acetylphosphate is converted by the acetate kinase (AK) to acetate and ATP. The citrate/acetate antiporter CitW releases the generated acetates to the surrounding environment and simultaneously imports the nutrient citrate (Kästner *et al.*, 2002).

The enzyme phosphotransacetylase (PTA) converts Acetyl-CoA and inorganic phosphate to free CoA and acetylphosphate. The latter is dephosphorylated by acetate kinase (AK) to form the two products acetate and ATP (Buss *et al.*, 2001, Iyer *et al.*, 2004). AK and PTA are present in aerobic or anaerobic microbes of the bacteria or archea kingdoms. They play a major role in the decomposition of organic material to methane. Under anaerobic conditions the overall reaction in bacteria generates energy in the form of ATP whereas archea use AK to activate acetate to acetylphosphate and decompose it to methane and carbon dioxide (Buss *et al.*, 2001). The overall process of citrate fermentation by *K. pneumoniae* produces 1 mol of ATP per mol of citrate.

D.1.2 The Sodium Citrate Symporter CitS of *Klebsiella Pneumoniae*

The citrate sodium symport protein CitS of *K. pneumoniae* (P31602, TC 2.A.24) consists of 446 amino acids with a molecular weight of 47588 Da. Analysis of the N- and C-terminal part of the transporter shows a sequence similarity between the two domains which are separated by a long loop between helix VI and VII. This generates an inverted membrane topology, which is present in other transport systems (Lolkema *et al.*, 2005). The number of predicted transmembrane helices depends on the algorithm used and varies between 9 and 12. Analysis by single site mutagenesis, alkaline phosphatase (PhoA) and biotin acceptor domain (BAD) fusion experiments, and labeling with cysteine specific reagents all revealed 11 transmembrane helices with two additional peptide segments, the so called pore-loops, which enter the membrane without crossing it (van Geest & Lolkema, 2000, 1999, Sobczak & Lolkema, 2005a).

The two loops Vb and Xa that penetrate the membrane are found between helices V and VI or X and XI. They play an autonomous role in the translocation mechanism and are important for functionality and specificity of the transporter (Figure 56). The mutation of N186V in the conserved region Vb affects the binding affinity of citrate to the transporter therefore indicating the involvement of that residue in the transport cycle (Kästner *et al.*, 2000). Fluorescence spectroscopy experiments of single site labeled C398 showed different accessibility to the solvent upon incubation with citrate (Kästner *et al.*, 2003) and, in addition, the loop Xa contains two cysteine residues which show a different accessibility depending on the catalytic state of the transporter (Sobczak & Lolkema, 2003).

Evidence for the presence of the pore loops in the class of secondary transporters is also found in the closely related glutamate transporter family (Sobczak & Lolkema, 2005b). Recent work on the sodium-coupled aspartate transporter Glt_{PH} of *Pyroccocus horikoshii* defines the importance of the two re-entrant hairpins: both are involved in substrate and ion binding (Yernool *et al.*, 2004, Boudker *et al.*, 2007).

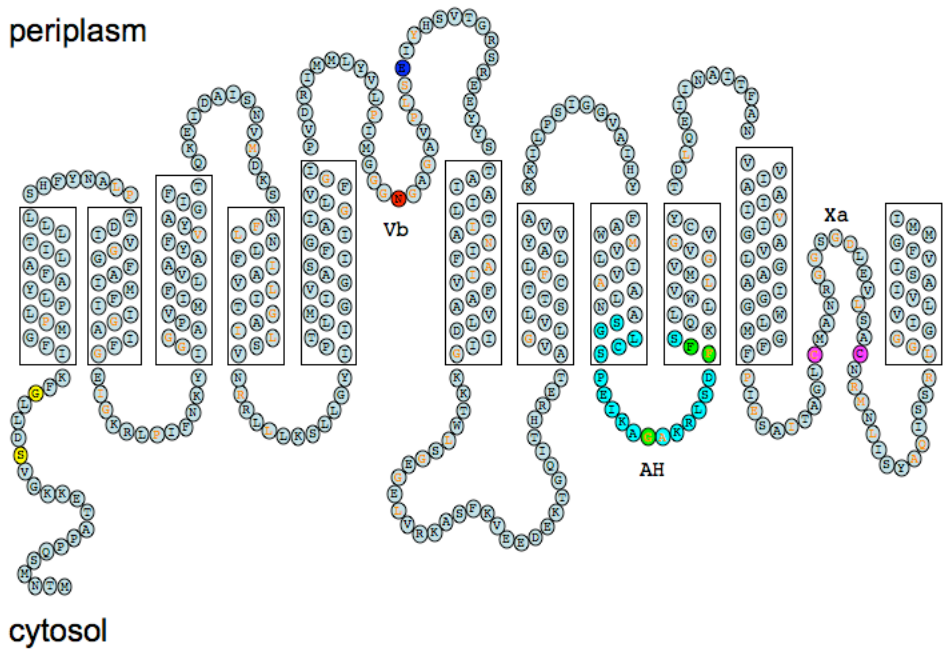

Figure 56: Topology of CitS. The eleven transmembrane segments and the loops penetrating the lipid bilayer (Vb and Xa) are shown. Transmembrane helices VIII and IX are connected by the amphipathic helix AH (shaded turquoise) that faces the protein with its hydrophobic side and contacts the water phase with its hydrophilic part. Amino acids marked in orange letters share > 80% identity with other Na^+-citrate or Na^+-malate transporters. A mutation N185V (shaded red) affected citrate transport whereas E196Q (shaded blue) did not change transport behavior (Kästner *et al.*, 2000). Cysteine mutants of the green shaded residues reduced the citrate uptake <20 % in membranes. The magenta shaded cysteines 398 and 414 are accessible from the periplasm (Sobczak & Lolkema, 2003). Transmembrane domains were predicted with HMMTOP (Tusnady & Simon, 2001).

Recently it was shown that the loop connecting helix VIII and IX resides in the cytosol and forms an amphipathic helix. The hydrophobic side of this helix faces the protein whilst the other face is hydrophilic and exposed to the water (Sobczak & Lolkema, 2005a). Since the accessibility depends on the catalytic state of the transporter, an alternating access transport model, comparable to the transport mechanism described for the glycerol-3-phosphate transporter of *E.coli* and the lactose permease of *E.coli,* can be assumed (Locher *et al.*, 2003, Abramson *et al.*, 2003, Huang *et al.*, 2003, Sobczak & Lolkema, 2005b).

As shown in Figure 55 the protein CitS of *K. pneumoniae* transports citrate across the membrane. Transport studies with whole cells or membranes showed that the transport process is Na^+ and ΔpH dependent and electroneutral (Dimroth & Thomer, 1990, Pos & Dimroth, 1996). The transported species are Na^+ and $Hcit^{2-}$ (Van der Rest *et al.*, 1991). Since the transport is also pH gradient dependent it was suggested that one Na^+ ion and a proton might be the counter ions during transport (van der Rest *et al.*, 1992). The cooperative effect of Na^+ ions during transport, however, suggested that two sodium ions are transported with one $Hcit^{2-}$ molecule (Lolkema *et al.*, 1994) and therefore the overall transport reaction was assumed to be:

$$Hcit^{2-}_{out} + 2Na^+_{out} + Na^+_{in}\,OH^-_{in} \rightarrow Hcit^{2-}_{in} + 2Na^+_{in} + Na^+_{out} + OH^-_{out}$$

with a net transport of $Hcit2^-$, one Na+ ion and one proton into the cell or, instead of the proton, a OH^-, which is transported out of the cell during the transport cycle. In addition, Pos & Dimroth (Pos & Dimroth, 1996) showed that the transporter contains a high and a low affinity binding site. The high affinity site of the external site is K_m = 140 µM whereas the internal low affinity site binds the substrate citrate with a K_m = 14 mM.

D.2 Results of the Sodium Citrate Symporter CitS

The aim of this project is the crystallization and structure determination of the first member of the 2HCT family that belongs to the class of secondary transporters. Protein production, construct design and purification strategies were carried out in a close and fruitful collaboration with Thomas Huber of the group of Andreas Plückthun, Institute of Biochemistry, University of Zürich.

Expression and purification of CitS was carried out as described in material and methods. The purification yielded ~ 2 mg purified protein per liter of cell culture. The protein was concentrated to 10-30 mg/l and initial screen of crystallization conditions were carried out according to the protocol described in material and methods. Improvement of the initial crystals obtained is described in the following chapters.

D.2.1 Improvement of Crystal Growth

The initial crystals that were obtained by the screening on a nano liter scale were tested by diffraction to prove the presence of protein. Although protein crystals were detected, their size was too small and their quality too poor for structural studies. Therefore the initial conditions had to be improved.

Figure 57: Initial crystals of CitS: The size of the black bar is 40 µm.

The initial conditions that were identified during screening (100 mM KCl and 15-25% PEG 400 or 50 mM Magnesium acetate and 15-25% PEG 400 both in a pH range of 6.5-8.4) were further examined and refined in 1+1 µl drops, using 24 sitting drop well plates. Crystals appeared after 1-3 days and grew to full size in 20-30 days. Rod like shapes, plates and triangles were observed and the biggest crystals reached dimensions of about 50x50x50 µm. Refinement included variations of different salts and their concentration (KBr, KI, NaCl, NaBr, CsCl, RbCl, $Mg(Acetate)_2$, KCl; 50-300 mM), the PEG concentration and composition (15-35%, PEG 200-400), as well as the different buffer systems: ADA, Tris or HEPES within their appropriate pH values. This fine tuning of the original conditions led to an increase in the diffraction power from 15 to about 8 Å.

To further improve the crystal quality, an additive screen including salts (mono or divalent), small organic molecules, sugars, amino acids, detergents, dyes or cryopreservatives, was performed in a 96-well plate format using 100 nl protein solution and 100 nl of reservoir solution containing 10 % of additive. Although the experiments were repeated with different starting conditions, none of the additives improved the crystal quality. Crystallization with 100 µM citrate in the reservoir solution prevented the formation of crystals. Additionally, several additives were manually added during crystallization in the µl format. These included bi- or tri-alcohols, lipid extracts from *E.coli* and different detergents to affect the original DDM detergent micelles and therefore potentially enable better crystal growth. A major improvement in the crystals was achieved by supplementing the reservoir solution with 3 µl of 10 % (w/v) of the detergent Cymal-7. Although the crystal shape and size were not tremendously affected, it improved the crystal handling since the crystals were less prone to stick to the bottom of the well. As with other additives used such as heptantriol, heptandiol or MPD, detergents such as glycosides, LDAO, DDAO, Mega-8 or Cymal-5 had no effect on crystal growth or crystal behavior.

Under the assumption, that the N-terminus of CitS is flexible which potentially prevents crystal contacts, new constructs with truncations at the N-terminus were made. The construct SD had a deletion of the first 15 amino acids from the original CitS sequence, whereas the construct GF started just before the first predicted α-helix. Additional truncations were introduced in the linker region between the N-terminal His-tag and the protein to generate the construct SH. The constructs were screened for initial crystallization conditions in the 96-well format and refined using the 24-well sitting drop plates.

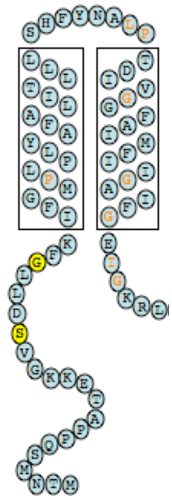

Figure 58: N-terminal truncations of CitS: The first 15 or 19 N-terminal residues were truncated generating the constructs SD or GF, respectively, both containing a 10xHis-tag attached to the N-terminus**.**

The construct SD crystallized under similar conditions as the original construct. Although the protein is shorter, the observed crystals did not diffract as well as those from the full length and the maximum resolution that could be obtained was 18 Å. In contrast, the construct GF crystallized in different conditions (10 mM sodium cacodylate, 100-200 mM NaCl, 8-13% PEG 4000, pH 5.5) and the shape of the crystals changed to a cubic form. Crystallization in these conditions was associated with a strong precipitation of the protein, but nevertheless small crystals could be obtained. However, the properties of these crystals were, compared to the original conditions, worse. Although multiple attempts were carried out, including seeding and buffer, salt and detergent exchange, the crystals size remained small, about 10 x 10 x 10 µm. Since the crystals diffracted only to 12 Å and cell parameters and space group were equal to the original crystal form, this construct and crystallization conditions were not considered for further experiments.

Deleting amino acids at the N-terminal end of the protein did not improve the quality of the crystals. However, an improvement was observed after the further deletion of residues in the linker region between the protein and the N-terminus. Diffraction quality was slightly enhanced compared with the original construct and all subsequent expression and crystallization trials were carried out with the construct SH that lacks the protease cleavage site between the His-tag and the protein.

Changing the detergent often enhances the properties of membrane protein crystals since the detergent belt around the protein favors or disturbs the intermolecular contacts. Therefore, the detergent DDM was exchanged, in turn, to various other detergents during purification. Maltoside detergents were superior with respect to aggregation compared to glucosides, Fos-cholines or LDAO, as deduced from initial experiments and size exclusion chromatography analysis. In these detergents, the protein eluted mainly as a single peak, while the “extremes” tridecyl- and nonyl- maltoside showed additional peaks (Figure 59).

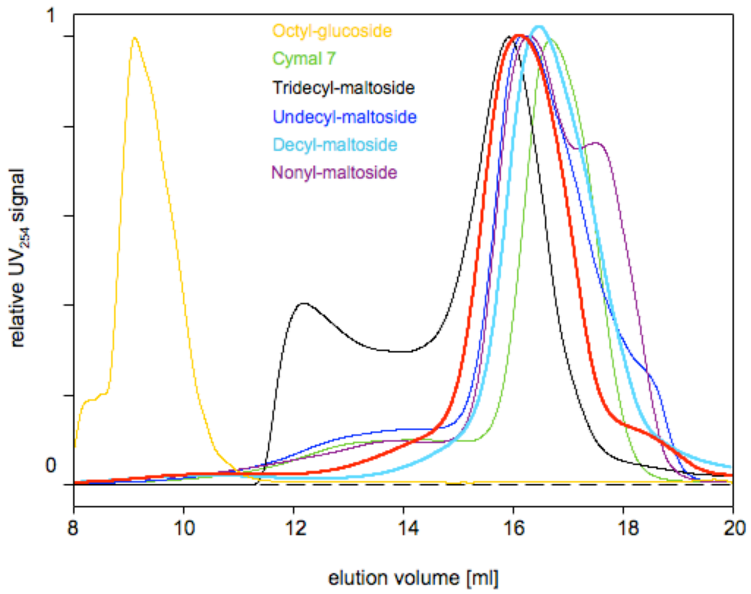

Figure 59: Behavior of CitS in the presence of different detergents. Elution diagrams of size exclusion chromatography experiments of CitS in the presence of different maltosides are shown. For comparison, the elution profile in presence of DDM is shown in red. Incompatible detergents, e.g. octyl-glycoside promoted aggregation and the protein eluted directly after the void volume of the column (Superose6 10/300 Pharmacia).

DDM was therefore exchanged to tridecyl β-D-maltopyranoside (TDM), undecyl β-D-maltopyranoside (UDM), decyl β-D-maltopyranoside (DM), nonyl β-D-maltopyranoside (NM) or cyclohexyl-heptyl β-D-maltopyranoside (CYMAL-7). Crystallization trials of CitS preparations in different detergent were set up using the 96-well and the 24-well plates. Crystals could be obtained from the preparations containing the detergents UDM, DM and TDM, but the diffraction quality of these crystals was not as good as of crystals in DDM.

To further improve the crystal quality, seeding experiments, including micro and macro seeding, were carried out. In macro seeding, a single crystal from a previous experiment is transferred to a new crystallization drop that already contains the protein and reservoir solution. Micro seeding is similar, although the seeds are much smaller and the crystals from a previous experiment are transferred differently into the new drop. Two transfer techniques are widely used: diluting the seeds in reservoir solution and subsequently adding this solution to the new experiment; or transfer using a cat whisker. For the latter, a the cat whisker is pulled through the original drop containing small crystals, and is then passed through the new drop, which contains the protein and reservoir solution.

Macro seeding was carried out by transferring crystals that had grown at the same pH, with a crystal mounting loop, to the new drop containing the reservoir and protein solution. Trials with higher precipitant concentration failed due to heavy precipitation of the protein and therefore the standard concentration of precipitants was used. Although the seeds started to dissolve initially, the final crystal size and shape was greatly enhanced by this technique. Similarly, micro seeding experiments led to a higher probability of crystal growth, resulting in a larger number of smaller crystals.

These modifications of the crystallization protocol in combination with post crystallization refinement eventually resulted in crystals with dimensions of ~ 300 x 100 x 100 µm (Figure 60) that diffracted in the best case to 3.6 Å.

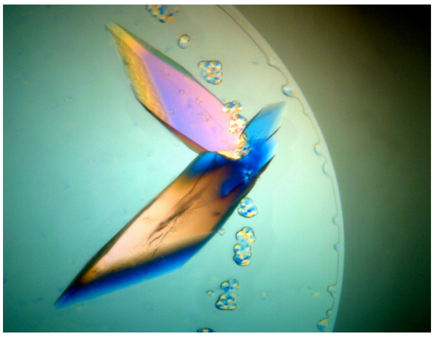

Figure 60: Crystals of CitS. Refined crystals viewed through polarized light. The field of view is 500 µm.

D.2.2 Post Crystallization Improvement of CitS Crystals

All crystallization experiments with CitS had to be done at 4°C and the crystals were sensitive to temperature changes. Crystals, therefore, required stabilization to enable data collection and flash freezing was mandatory since diffraction data of CitS can only be obtained using synchrotron radiation. Freezing is standard today, especially for delicate crystals that suffer dramatically from radiation damage. The disadvantage of freezing the crystal is the additional manipulation step of cryo soaking and the freezing process itself. Both can introduce disorder in the crystal that leads to a higher mosaicity and potentially lower the diffraction quality of the crystal. To freeze crystals, suitable cryo-preservative conditions must be found and the freezing protocol itself considered. Cryo protection agents, for example glycerol, ethylene glycol or low molecular weight PEGs, prevent the water molecules from forming ice during cooling, leading to two positive effects: first, the formation of ice could destroy the crystal lattice, especially if the protein crystals, as in the case of membrane proteins, have high solvent content. Secondly, the formation of ice generates strong background diffraction during data collection.

An ideal situation occurs if the condition in which the crystals have grown already promotes the formation of vitreous water and prevents the formation of ice. Therefore, the identified crystallization conditions of CitS were adjusted to contain enough glycerol or ethylene glycol, two well known cryoprotectants, to enable the glass formation during freezing. Unfortunately, the substitution of 15% water by glycerol or ethylene glycol in the reservoir solution prevented the formation of crystals and, since at least 24% of either glycerol or ethylene glycol are required, this technique could not be used. The crystals were therefore transferred to solutions containing the original reservoir solution in which glycerol, ethylene glycol or PEG 400 replaced the water at a concentration of 24, 24 or 32.5% respectively. When using glycerol or ethylene glycol as a cryoprotectant, the crystals were sequentially dipped into solutions containing 5, 10, 15, 20 or 24% of the cryo preservative, to avoid dramatic changes in the environment of the crystal. When using PEG 400 as a cryoprotectant, the concentration was slowly increased in steps of 2%. The diffraction data analysis clearly indicated that ethylene glycol was better than glycerol, and equally good as PEG 400, although, with the latter, it was necessary to increase the soaking time from ~5 to a minimum of 20 seconds. Crystals soaked with PEG 400 as the cryoprotectant showed the lowest mosaicity, but the resolution limit of the diffraction data was slightly better with ethylene glycol and therefore ethylene glycol was mainly used as the cryo preservative.

Liquid nitrogen, liquid propane and gaseous nitrogen at 100° K promote different velocities in which the crystals reach the glass transition temperature of 136°K for pure water and therefore the three possibilities were examined. All three could be used to obtain ice-free crystals of CitS, although the cryo buffers for liquid propane had to contain 5% more ethylene glycol. In addition, the crystals that were frozen in liquid propane showed a higher background during measurement since the propane was prone to stick to the crystals. Freezing of the crystals in the gaseous nitrogen stream was possible, but the direct freezing in liquid nitrogen at 4°C was more reliable and reproducible. Therefore, this method was used to freeze the CitS crystals.

No other post crystallization treatments of the crystals, including several steps of dehydration and rehydration with different concentrations of PEG or humidified gaseous streams improved the crystal quality. Similarly, cross-linking experiments of the crystals with glutaraldehyde, performed according to Lusty et al (Lusty, 1999), did not improve handling or diffraction properties of the crystals.

A summary of the improvement of the crystallization process, including all of the above mentioned steps, is illustrated in Figure 61. The image shows the diffraction produced by various crystals at different stages during the crystallization refinement process. Figure 61 A shows a diffraction pattern with spots to a resolution of 12 Å of crystals that were obtained during initial screening. Crystallization in 24-well plates, combined adjustment of the precipitant PEG 400 resulted in the first complete data set collected with a resolution of 8 Å (Figure 61 B). Macro seeding and post crystallization refinement improved crystal size and quality, and a diffraction pattern, of routinely obtained crystals, with spots up to 4 Å is illustrated in Figure 61 C. Nevertheless, the individual crystals have tremendous variations in their diffraction quality and more than 100 crystals had to be screened to obtain a complete data set to 3.6 Å (Figure 61 D).

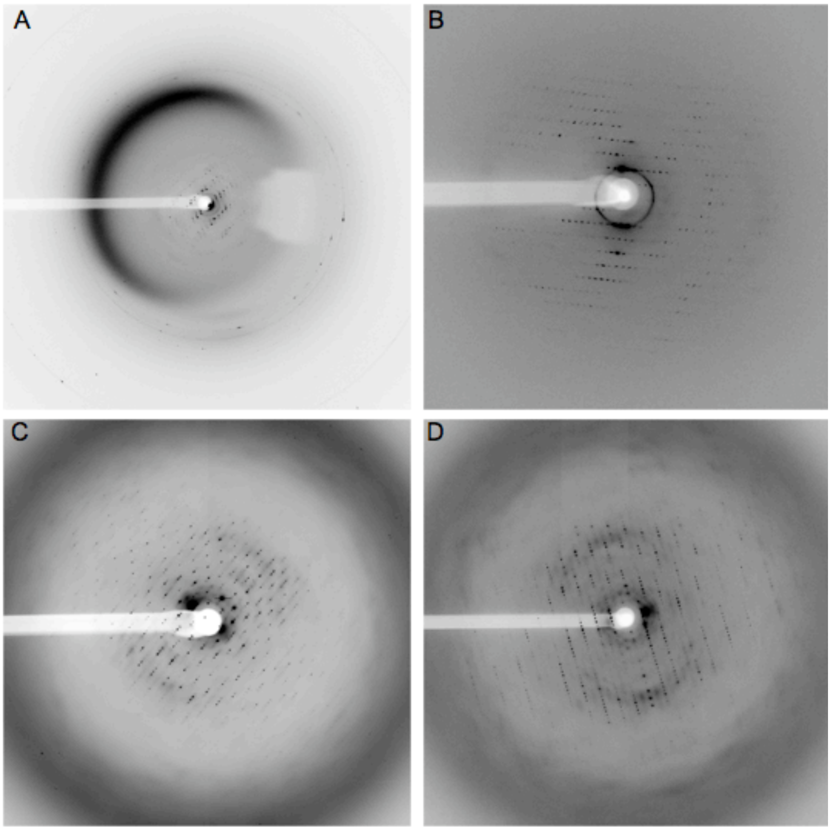

Figure 61: Diffraction images at different stages of crystallization refinement: A: diffraction of crystals obtained in the initial crystallization trials ~12 Å; B: after refinement of the precipitant PEG 400 the first complete data set to a resolution of ~8 Å was recorded; C: diffraction pattern routinely obtained of the refined CitS crystals, ~4 Å; D: Diffraction image of a crystal of CitS obtained after screening more than 100 crystals resulted in a resolution of 3.6 Å.

D.2.3 Crystallographic Analysis

To obtain the highest quality data and resolution, data sets were measured at cryogenic temperatures of 100° K at the Swiss Light Source, Villigen, Switzerland.

The CitS crystals diffracted to a resolution > 4 Å using a synchrotron light source, but the diffraction power of the crystals was dependent upon the orientation of the crystal. All measured crystals were highly anisotropic with the highest resolution of ~4 Å in one direction but only 5-6 Å ~90° away. Anisotropic diffraction is common in membrane protein crystals and is mainly due to the packing properties of the molecules. In addition to their anisotropic behavior, the crystals of CitS were sensitive to the strong radiation of the synchrotron beamlines. In general, protein crystals are sensitive to the X-ray beam, since the high energy produces radicals which diffuse through the crystal and thus destroying the uniformity of the crystal lattice. Figure 62 shows an example of the diffraction quality of a crystal and the loss of diffraction power during data collection. In Figure 62A, the last image of the data set, which diffracts to 7 Å, is shown. The same ϕ angle after translation is shown in Figure 62B in which spots to 4 Å are visible by eye on the image. To overcome the radiation damage, the crystals were therefore translated during the data collection and the multiple data sets obtained were scaled together after individual integration. Since radiation damage mainly depends on the total dose a crystal absorbs (Murray *et al.*, 2005, Ravelli & Garman, 2006), it was also possible to obtain complete data sets with an attenuated beam. The beam intensity was reduced by 50 or 70% by inserting aluminum filters into the pathway of the incident X-ray beam. The attenuation of the beam had no effect on the resolution of the measured crystals.

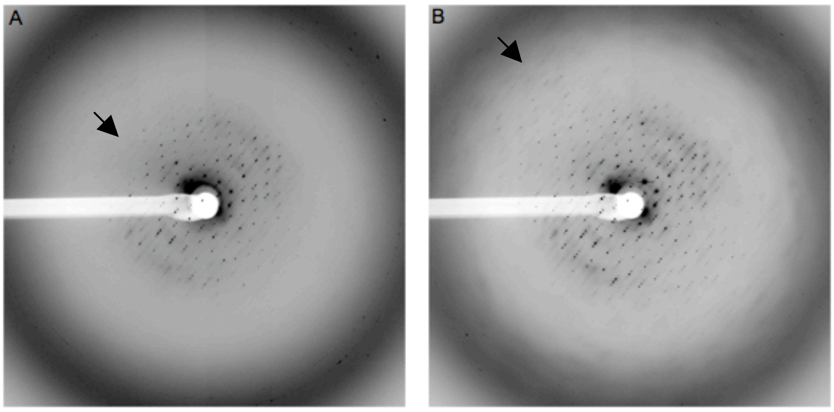

Figure 62: Radiation damage of CitS crystals: A: diffraction image of a crystal after exposure of 180 frames. B: Diffraction image of the same oscillation range after translation. Arrows indicate visual diffraction spots at 7 respectively 4 Å.

The homodimer CitS consists of two molecules of the protein CitS with a total molecular weight of 95.2 kDa. With a cell volume of 1957693 Å^3, the most probable number of homodimers in the asymmetric unit is 2 or 3, with a solvent content of 76.1 or 64.2 %, respectively (Table 6). The most frequently observed values of the Matthew parameter are in the range of V_M = 2.0 to 2.6, but these numbers are determined for soluble proteins. In the case of membrane proteins, where the detergent around the molecule must be taken into account, the values of V_M are higher and solvent contents of 70 % (V_M = 4.7) are common. Static light scattering analysis of the CitS protein in a detergent buffer containing DDM suggested that CitS is a homodimer in solution. The protein was surrounded by detergent molecules with a weight of ~100 kDa and the entire complex showed a molecular weight of ~200 kDa (T. Huber, personal communication). According to this molecular weight, the most probable number of complexes in the asymmetric unit is 2, with a V_M of 2.45 and a solvent content of 49.8%. Note, that these numbers are calculated with a molecular weight of the complex of 200 kDa, and the correct value of VM assuming 100 kDa of protein is 5.15 with a solvent content of 76%.

Table 6: Cell Content Analysis P2

Homodimers in the asymmetric unit	Matthews coefficient V_M	Solvent content (%)
1	10.29	88.06
2	5.15	76.11
3	3.43	64.17
4	2.57	52.22
5	2.06	40.28

To further characterize the cell content of the crystals, a self-rotation function and native Patterson map were calculated (Collaborative Computational Project Number 4, 1994). A section of the native Patterson map where additional peaks would indicate the presence of molecules related by a translation is shown in Figure 63. The map was calculated using data in range of 15-4 Å and a peak threshold > 4 σ. Apart from at the origin, no peaks were observed in the different sections of the map. Therefore a pure translational symmetry between the molecules can be ruled out. A plot of the self-rotation function, which detects the relative orientation of the molecules independent of their position in the asymmetric unit, is shown in Figure 64. Beside the two peaks at ω =90° and ϕ = ±90°, which are generated by the two-fold screw axis present in space group $P2_1$, four additional peaks (I - IV in Figure 64) at ω=14° ϕ =0°, ω= 45° ϕ = 104°, ω=67° ϕ =94° and ω= 28° ϕ = 118° are found in the κ =180° section with a height of 74%, 66%, 55% and 45% of the origin peaks, respectively. Peak I, which is located close to the pole, is also present in the κ= 90° section at ω = 76° and ϕ = 0° indicating a tetrameric (dimer of a dimer) CitS in the crystal. The four-fold axis of this tetramer would be nearly parallel to the a-axis of the crystal. Peak II could therefore be explained as the diagonal through the tetramer. The weaker peaks III and IV might correspond to a second tetramer in the asymmetric unit, which is again parallel to the a-axis but slightly tilted ~20°, compared to the first tetramer. Reducing the integration radius to a value of 20 Å resulted in less defined peaks and the peaks III and IV were not appearing at all. This might result from the similarity of the rotation angle of the two protomers comprising the homodimer, thus indicating the approximate size of the monomer to be approximately 20 Å.

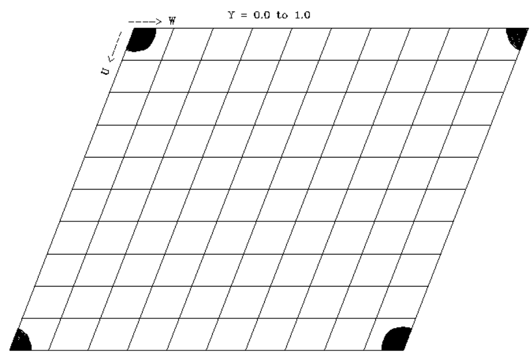

Figure 63: Native Patterson map. No peaks were detected above 4σ, indicating the absence of translational symmetry.

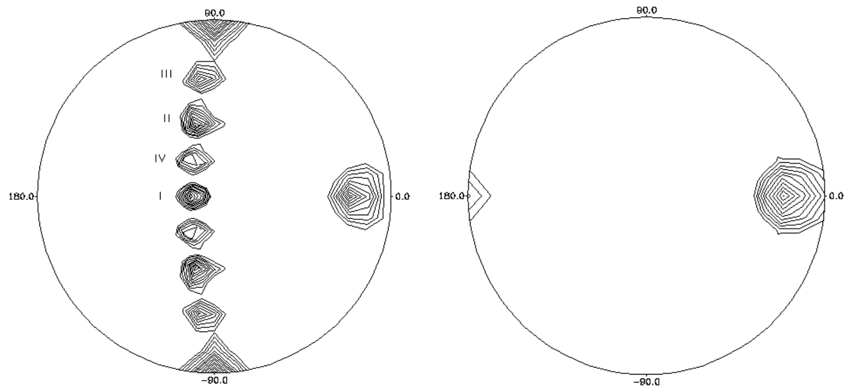

Figure 64: Self-rotation function. The plot was generated using the program POLARRFN of the CCP4 program suite (Collaborative Computational Project Number 4, 1994). Data was used in the range of 20-7 Å with an integration radius of 25 Å; a smoothing factor of 6 Å and sharpening was applied by setting the B-value to -20. Peaks were contoured at >30% of the origin peak in steps of 5%. The κ = 180° and the κ = 90° sections are shown on the left and right, respectively. Note that changing the value of the integration radius between 25-40 Å had no effect on the appearance of the plot.

D.2.4 Heavy Atom Derivative Preparation of CitS Crystals

Since the calculation of an electron density map depends on the availability of the intensities and the corresponding phase angles of each reflection (**F**(hkl)=|F|exp[iα]) and since structures of homologues of CitS are not known, molecular replacement could not be used to calculate initial phases. Therefore, SAD, MAD, MIR, SIR and SIRAS techniques have to be considered to solve the protein structure. These methods require that crystals contain either heavy atoms or selenomethionine. The anomalous signal from the sulfur atoms present in methionines and cysteines in the protein can also be used to phase the protein structure but the diffraction quality of the CitS crystals is far too low to exploit this weak signal for structure determination.

D.2.4.1 Isomorphous Derivatives using Heavy Metals

Isomorphous derivatives can be prepared by adding the heavy metals during crystallization or, by soaking preformed crystals in solutions containing the heavy metals. The first procedure is more gentle and thus preferred, although the addition of an extra compound to the reservoir solution can prevent crystallization completely. Two initial incomplete factorial screens in the 96-well format were set up by adding 16 different heavy metal solutions with a final concentration of 100 µM to the crystallization conditions of CitS (50 mM Tris pH 7.0-8.5, 100 mM KCl and 20-21.5% PEG400, Figure 65). In 10 of the 16 conditions tested, nice crystals were obtained and in an additional 4 conditions crystalline material appeared. The two mercury derivatives thiomersal and methyl mercury (I) chloride prevented crystal growth. Successful heavy metal compounds were then applied to further crystallization experiments using 24-well plates. Trials with the two mercury compounds that failed in the first round of screening were set up at a ten times lower concentration and crystals were obtained.

In addition to the co-crystallization experiments, various soaking experiments with pre-existing crystals were carried out. Heavy metals (100 µM stock solution in water) were diluted to a final concentration of 0.1-2 mM in the appropriate reservoir solution and CitS crystals were added. The crystals were harvested at different time points (30 s, 2 min, 15 min, 30 min, 1h, 2h or 12 hours), cryo protected and frozen in liquid nitrogen. Possible derivatives were analyzed by diffraction and a fluorescence spectrum. Properties of derivatives that provided complete data sets are summarized in the appendix in Table 9.

Data sets of possible derivatives were recorded at appropriate wavelengths keeping Friedel`s pairs separate to detect a possible anomalous signal present in the data. The mean □value assuming Friedel`s law as true, divided by the mean ọvalue assuming Friedel`s law as false gives an estimation of the presence of anomalous scatterers in the crystal. Values > 1.00 indicate the contribution of anomalous scattering to the intensities and these data sets were further analyzed to find the positions of possible heavy atom sites.

pH 7.0	$K_2Pt(CN)_4$	Dichloro (2,2	$OsCl_3$	Thiomersal	Pb(NO3)2	$UO_2(OCOCH_3$	$KAu(CN)_2$	$HoCl_3$
PEG 400 20%	A1	A2	A3	A4			A5	A6
PEG 400 20.5%	A7	A8			A9	A10	A11	A12
PEG 400 21%	B1	B2	B3	B4	B5	B6		
PEG 400 21.5%			B7	B8	B9	B10	B11	B12
pH 7.5								
PEG 400 20%	C1	C2	C3	C4		C5	C6	
PEG 400 20.5%	C7	C8	C9		C10	C11		C12
PEG 400 21%	D1			D2	D3	D4	D5	D6
PEG 400 21.5%		D7	D8	D9	D10		D11	D12
pH 8.0								
PEG 400 20%	E1	E2	E3		E4	E5		E6
PEG 400 20.5%	E7		E8	E9	E10		E11	E12
PEG 400 21%		F1		F2	F3	F4	F5	F6
PEG 400 21.5%	F7	F8	F9	F10		F11	F12	
pH 8.5								
PEG 400 20%		G1	G2	G3	G4	G5		G6
PEG 400 20.5%	G7		G8	G9	G10	G11	G12	
PEG 400 21%	H1	H2	H3			H4	H5	H6
PEG 400 21.5%	H7	H8		H9	H10		H11	H12

pH 7.0	Hg $(OAC)_2$	TaBr	$(NH_4)_2WS_4$	K_2PtCl_6	$KAuCl_4$	$Hg(CH_3)Cl$	Hg_2Cl_2	TlCl
PEG 400 20%	A1	A2	A3	A4			A5	A6
PEG 400 20.5%	A7	A8			A9	A10	A11	A12
PEG 400 21%	B1	B2	B3	B4	B5	B6		
PEG 400 21.5%			B7	B8	B9	B10	B11	B12
pH 7.5								
PEG 400 20%	C1	C2	C3	C4		C5	C6	
PEG 400 20.5%	C7	C8	C9		C10	C11		C12
PEG 400 21%	D1			D2	D3	D4	D5	D6
PEG 400 21.5%		D7	D8	D9	D10		D11	D12
pH 8.0								
PEG 400 20%	E1	E2	E3		E4	E5		E6
PEG 400 20.5%	E7		E8	E9	E10		E11	E12
PEG 400 21%		F1		F2	F3	F4	F5	F6
PEG 400 21.5%	F7	F8	F9	F10		F11	F12	
pH 8.5								
PEG 400 20%		G1	G2	G3	G4	G5		G6
PEG 400 20.5%	G7		G8	G9	G10	G11	G12	
PEG 400 21%	H1	H2	H3			H4	H5	H6
PEG 400 21.5%	H7	H8		H9	H10		H11	H12

Figure 65: Incomplete Factorial Screen. 16 different heavy metal solutions (potassium platinate (II) cyanate, dichloro (2,2,6`,2``-terpyridine)platinium (II), osmium (III)chloride, lead (II) nitrate, uranylacetate, potassium di-cyanoaurate (I), holmium (III) chloride, mercury (II)acetate, thaliumbromide, di-ammonium tetrathiotungstate (VI), potassium hexachloro platinate (IV), potassium tetra-chloro aurate (III), methyl mercury (I)chloride, thalium (I) chloride) were added to a final concentration of 100 µM to the crystallization buffer. A1-H12 indicates the number of the wells and colors green, yellow and red correspondent to observed and nice crystals, crystals or no crystals at all, respectively.

Positions of the heavy atoms were searched individually with the programs SHELX97, CNS and phenix.hyss (Schneider & Sheldrick, 2002, Brünger *et al.*, 1998, Adams *et al.*, 2004). The resulting positions of the heavy metals were used to calculate a Patterson map that was compared to the anomalous Patterson map obtained from the data. In addition, data sets that showed equal cell parameters to the native data were analyzed by difference Patterson maps to obtain the positions of the heavy metals. Unfortunately, no heavy metal positions could be deduced from the various data sets collected.

D.2.4.2 Derivatization with L-selenomethionine

An alternative to soaking crystals with heavy atoms is the incorporation of heavy scatterers to the protein at the stage of transcription. The amino acid L-methionine is replaced by L-selenomethionines, which contain a selenium atom at the position of the sulfur. The presence of these modified residues in the protein sequence, and thus in the crystal, allows the selenium atoms to be used to determine initial phases of the reflections.

Selenomethionine derivatized CitS was prepared and purified according to the above-mentioned protocol, except that the cells were grown in minimal media, supplied with all amino acids and selenomethionine (per liter media: 5.4 g K_2HPO_4, 1.6 g KH_2PO_4, 0.8 g $(NH_4)_2SO_4$, 0.2 g Na-citrate, 80 mg MgSO4•7 H_2O, 32 mg thiamine, 32 mg thymine, 100 mg of each amino acid including selenomethionine and 16 ml 40% (w/v) glucose. The incorporation of selenomethionines was verified by MALDI-MS spectra. Crystals were obtained but they were much smaller and grew to a final size of ~20 x 20 x 20 μm^3. Further screening of the crystallization condition with the SeMet derivatives showed that further addition of 1 mM holmium (III) chloride promoted larger crystals of approximately ~50 x 30 x 30 μm^3. These crystals diffracted about to 6 Å and a complete data set to 7 Å could be obtained. Although the anomalous signal was present and complete MAD data sets were recorded, the low resolution and the high number of selenomethionines in the unit cell (assuming 4 CitS monomers in the AU: 4 x 17 x 2= 136) prevented structure determination of the transporter.

D.3 Crystallization of CitS - Complexes

The crystallization of the membrane protein CitS with two types of proteinaceous binding partners was carried out. The aim was to grow crystals diffracting to higher resolution than crystals of CitS alone. First attempts involved the crystallization of CitS in complex with Fab fragments while the second trials used the recently developed DARPin molecules as binding partners.

D.3.1 Complex Crystallization of CitS with Fab Fragments

Four Fab fragments were selected from the HuCAL GOLD® library (Knappik et al., 2000) and all four showed high binding affinities in the nano molar range (Röthlisberger *et al.*, 2004). Individual complexes for each of the fragments with CitS were prepared and crystallized. The preparation included the following steps: separate expression and purification of both proteins, mixing in a molar ration of CitS : Fab =1 : 3 and subsequently, a size exclusion chromatography purification step to separate the complex and the unbound antibody fragment. As yet, no crystals of the complexes have been obtained. Nevertheless, the control experiment with the Fab f3p4 alone produced crystals that diffracted to 1.9 Å. Structure determination and refinement of this structure are described in Chapter E.

D.3.2 Complexes of CitS with selected DARPin Proteins

To further increase the probability to obtain better diffracting crystals, designed ankyrin repeat proteins (DARPins) as binders were created. The molecules were selected from an existing library (Binz *et al.*, 2004) by ribosome display and phage display against the native CitS protein (Huber *et al.*, 2007). DARPin molecules that showed positive binding signals in ELISA experiments were further analyzed by size exclusion chromatography. A typical size exclusion chromatography experiment (Superose6, 10/300, Pharmacia) is shown in Figure 66. The dashed lines show CitS, which elutes at ~12.2 ml and a DARPin, which elutes at ~16 ml. The solid line shows the chromatogram of the complex which shows a shift of ~0.5 ml towards higher molecular mass, indicating the presence of a CitS DARPin complex. Excess DARPin could be separated from the complex, which eluted at ~16 ml.

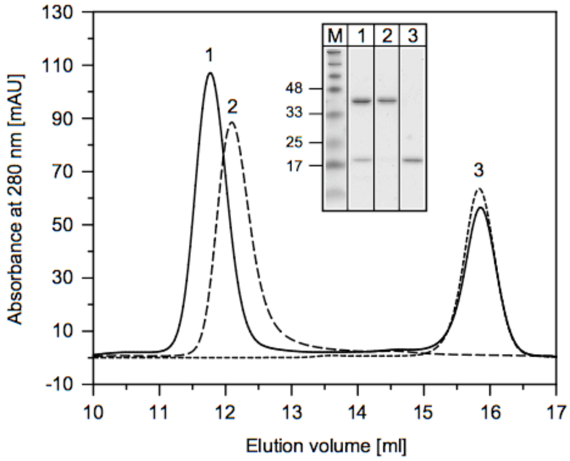

Figure 66: Size exclusion chromatography of the CitS-DARPin complex. Size exclusion chromatography analysis of both proteins alone and the CitS DARPin complex are shown as dashed and solid lines, respectively. The insertion shows the SDS-PAGE analysis of the individual peaks**.**

D.3.2.1 Crystallization of the CitS-DARPin Complexes

Proteins for co-crystallization experiments were purified separately and mixed with an excess of DARPin. CitS was purified using the mentioned protocol. DARPins were purified according to the established protocol (T.Huber, personal communication). Excess of DARPin was removed by size exclusion chromatography chromatography.

Crystallization trials were set up with selected DARPins that showed a clear shift in size exclusion chromatography analysis. The complex was concentrated to ~25 mg/ml (Amicon Ultra, 50 kDa cutoff, Millipore) and initial crystallization conditions were screened in the 96-well plate format. Initial conditions were refined using the 24-well format resulting in the final conditions containing 1 µl protein solution mixed at 4°C with 1 µl of 50 mM MES (pH 6.0-6.7) or 50 mM HEPES (pH 6.8-7.2), 150 -250 mM $CaCl_2$ and 25-28% PEG400. Crystals appeared after 3-5 days and grew to full size in ~2 weeks (Figure 67). Two types of crystals could be observed: rhombohedral like plates that were surrounded by a protein skin were located at the surface of the drop and rod like crystals, predominantly at the bottom of the drop, those were mostly attached to the plastic surface.

The presence of the complex in the crystals was verified by SDS-PAGE analysis of dissolved crystals. Therefore, crystals were washed 6 times in reservoir solution, dissolved in pure water and applied to the SDS-PAGE (Figure 68). Both proteins were visible and showed the expected apparent molecular weights of 35 kDa for CitS and 17 kDa for cp34h_15.

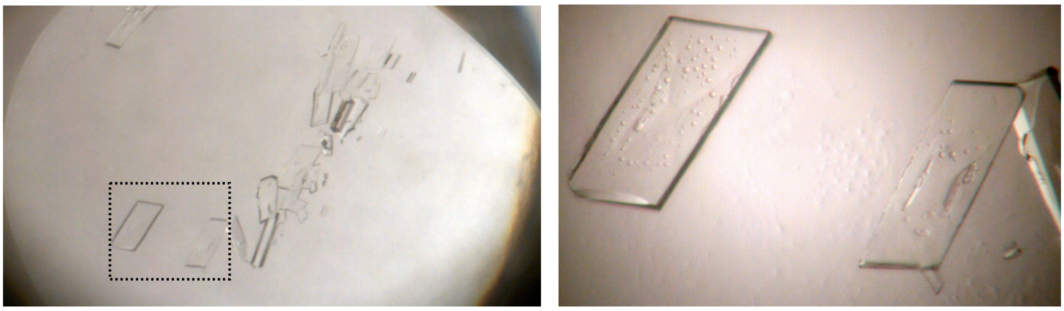

Figure 67: CitS-cp34h_15 complex crystals. Crystals obtained from a co-crystallization experiment of CitS and the DARPin cp34h_15. The right image shows a zoomed view of the rectangle in the left image. The lower left crystal has an approximately size of 280 x 140 x 30 µm^3.

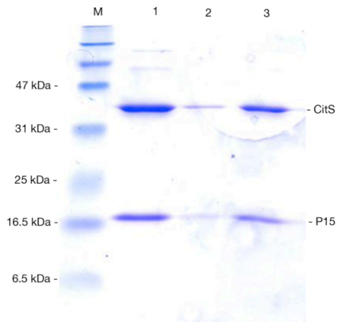

Figure 68: SDS-PAGE analysis of complex crystals. Different amounts of washed and dissolved crystals were applied to lanes 1-3. Lane 3 contained the most extensively washed crystals. Both, CitS and the DARPin cp34h_15 at apparent molecular masses of 35 kDa and 17 kDa, respectively, were visible, verifying the presence of the complex in the crystals.

As with with the non-complexed CitS, additional experiments to improve the crystals were set up. These included detergent exchange during protein purification and stabilization procedures of the crystals formed as well as testing freezing and cryo-conditions. Detergent exchange with shorter alkyl maltoside chains led to a similar crystallization condition, although the optimal pH of crystal growth slightly decreased to pH 6.5 for UDM and pH 6.0 for NM. However, both the size and the appearance of the crystals were worse than those obtained from the crystals grown in DDM. The optimal cryo solution contained the original reservoir components with the water changed to 35 %

PEG 400 or 22% glycerol. To prevent sudden changes in the crystal environment, the concentration of the cryoprotectant was slowly increased by 2 or 5% for PEG 400 or glycerol, respectively. Cryo-protected crystals were frozen in liquid nitrogen.

D.3.2.2 Crystallographic Analysis of CitS-DARPin Complexes

Two data sets of the complex crystals were collected and both inherited a high anisotropy in the data as seen in Figure 69. Similar to the native CitS crystals, a high resolution area of ~4 Å in one direction could be recorded but only 5-6 Å ~90° away.

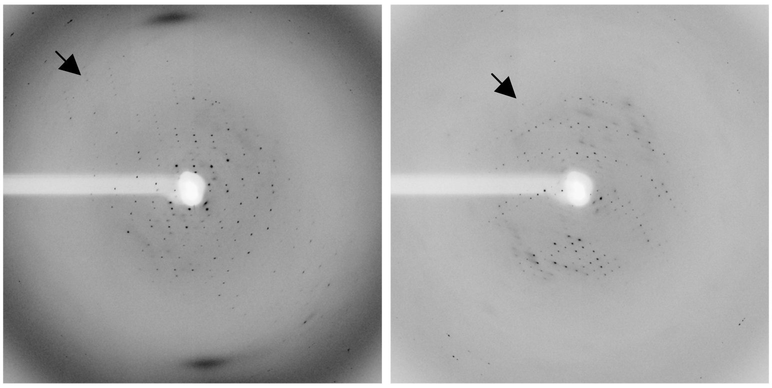

Figure 69: Anisotropic behavior of CitS-DARPin complex crystals. Diffraction images of complex crystals at 0 and 90° are shown. The left image shows the high resolution area with spots visible up to 4 Å. The right image illustrates the diffraction 90° away. In contrast to 0° the resolution of the crystal is decreased to 6 Å.

The data obtained from the crystal grown in DDM showed a resolution limit of 4.5 Å and a full data set could be obtained to 5 Å. In contrast, crystals grown in UDM diffracted to 4 Å and a complete data set could be obtained to 4.5 Å. Both data sets were in the same triclinic space group but had different cell parameters of a = 91.16 Å, b= 103.13 Å, c = 126.78 Å, α = 111.46°, β = 95.07° and γ = 94.01° and a = 86.62 Å, b = 98.64 Å, c = 121.34 Å, α = 70.95°, β = 81.30° and γ = 90.34°, respectively. The κ = 90° section (not shown) shows no peaks above the threshold, indicating that the DARPin molecules bound disable the formation of the tetramer, which was observed in the crystals of CitS alone (Figure 64). Two additional twofold axis were detected along ϕ = 127° and ω = 65° and ϕ = 47° and ω = 118°.

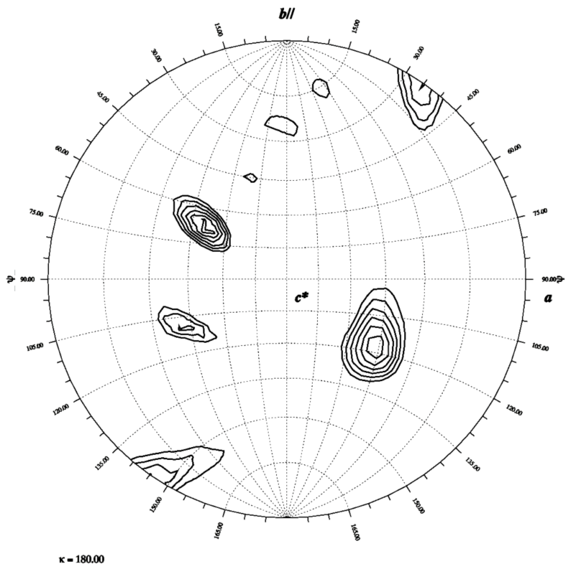

Figure 70: Self-rotation function of complex crystals. The plot of the k=180° section was generated using the program GLRF (Tong & Rossmann, 1997) with data in the range of 15-5 Å. Peaks are countered > 45 % of the origin peak height in steps of 5%. Two additional two fold axes are located at ϕ = 127° and ω= 65° and ϕ = 47° ω = 118°.

In contrast to the data obtained from non-complex crystals, the Patterson map of the data collected from the CitS-DARPin crystals showed an additional peak. The peak had a height of ~11% of the origin, and was located at the fractional positions u= 0.8843, v= 0.7384 and w= 0.4164 (equal to u= 0.1157, v= 0.2616 and w= 0.5836), indicating a possible pseudo translational symmetry between the molecules (Figure 71).

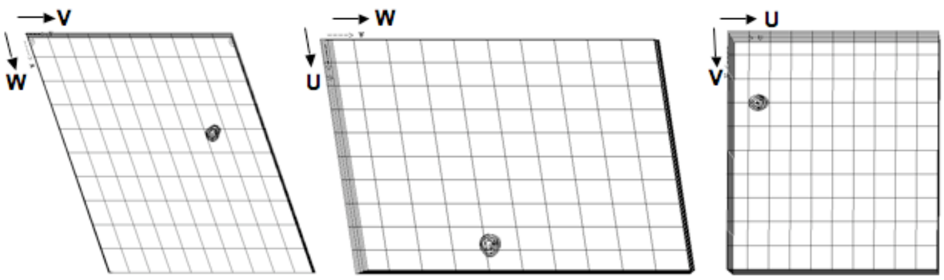

Figure 71: Plot of the Patterson map of the complex crystals. Section of u, v and w are shown from the left to the right. The map was calculated in CCP4 (Collaborative Computational Project Number 4, 1994)using data in the range of 30-4Å. Peaks are contoured < 3 σ in intervals of 0.5 σ.

The cell content analysis of the P1 crystal form, assuming a pure protein content of a homodimeric CitS and 2 bound DARPins with a molecular mass of the complex of ~130 kDa, indicates the presence of 2 complex molecules in the asymmetric unit (Table 7) with a V_M of 3.72 and a solvent content of 67 %. The results of the self rotation analysis and the cell content analysis indicate the presence of two distinct molecules in the cell that are related by a two fold symmetry. Each complex molecule itself contains a two fold symmetry between each of the protomers.

Table 7: Cell Content Analysis P1

Complexes in the asymmetric unit	Matthews coefficient V_M	Solvent content (%)
1	7.44	83.47
2	3.72	66.95
3	2.48	50.42
4	1.86	33.90

D.4 Discussion and Conclusion

This chapter has described the successful purification, crystallization and preliminary crystallographic analysis of the first member of the 2HCT family of membrane proteins. The protein CitS was expressed in a functional form and sufficient protein of high quality and homogeneity was obtained. The amount of purified protein was in the range of 1.5 – 2 mg purified CitS per liter of culture and, therefore, the significant hurdle in structural studies of membrane proteins, the low expression and purification yields, was overcome. Size exclusion chromatography analysis in several detergents revealed a single protein species with a molecular mass of ~ 200 kDa (Figure 59). The mass corresponded to the dimer assembly of the transporter reported by Pos (Pos & Dimroth, 1996). The amount of detergent around the transporter dimer is high and was estimated to account for half of the mass of the analyzed species.

Crystals of the protein were obtained and their diffraction properties improved to obtain data to 3.6 Å. The main improvement was achieved by seeding the crystals. The crystals grew to a larger size and the reproducibility was enhanced, thus enabling the search for derivatives for phasing. The transporter crystallized in the monoclinic Laue class P2 with apparently 4 or 8 molecules in the asymmetric unit, generating a Matthews Parameter of 5.2 and 2.6 and a solvent content of 76 and 52 %, respectively. In contrast to soluble proteins, for which the Matthews parameter allows a good approximation of the number of protein molecules in the unit cell, the number of molecules of membrane proteins is less predictable, since both protein and detergent molecules are present in the unit cell. Assuming a detergent to protein ratio of 1 : 1, as concluded from static light scattering analysis, the most favorable number of molecules in the asymmetric unit would be 4. Since the number of detergent molecules in the crystallized material is unknown, the corresponding number of protein subunits cannot be estimated correctly and a number of up to 8 molecules would be feasible. The tetramer, present in the self-rotation function and the Matthews coefficient analysis, would therefore correspond to 4 or 8 molecules in the asymmetric unit cell of the crystals.

A high number of molecules in the cell has advantages and disadvantages. At the resolution obtained for these crystals, non crystallographic symmetry and averaging of the individual chains often leads to improved electron density maps and therefore would help in model building. On the other hand, the high copy number of protein chains in the cell disabled the structure solution with selenomethionine derivative crystals. The crystals obtained from protein that was expressed in the presence of selenomethionine, were smaller and did not diffract as well, but nevertheless, data to 7 Å were obtained. Positioning of the selenium sites could not be achieved since the high numbers of molecules in the unit cell, as mentioned above, would generate ~ 130 selenium sites. Even at a higher resolution of ~ 2 Å or better, this enormous number of scattering atoms is unfavorable for solving the structure of any protein. Unfortunately no experiments including truncation of the protein, detergent and buffer exchange resulted in a different crystal form and with the high copy number of molecules structure determination was hindered. The resolution of the crystals soaked or co-crystallized with heavy atoms was low, ranging between 5 and 8 Å. Therefore, it was not possible to determine the positions of the heavy atoms. It was expected, that the soaks in particular with mercury compounds, provide some phasing information, since at least two of the 5 cysteines have been reported to be chemically accessible (Kästner *et al.*, 2003, Sobczak & Lolkema, 2003, 2004), and could react with mercury compounds. However, despite testing several mercury compounds, a clear signal in a Patterson map could not be observed.

As an alternative to the derivatization of crystals, other techniques were considered to enable structure determination. Crystal of the complex of CitS with antibody fragments could not be produced by co-crystallization experiments. There are two possible explanations for the lack of crystals (i) the low solubility of the Fab f3p4, insoluble above 12 mg/ml; (ii) the presence of the membrane protein in the complex in a different conformation rendering it unsuitable for crystallization. The latter was observed during crystallization experiments in the presence of citrate. If the transporter substrate was present in crystallization experiments, no crystals appeared and, significantly, crystals grown in the absence of citrate dissolved rapidly after incubation with the substrate.

In contrast to the co-crystallization experiments with the antibody fragments, the ankyrin-CitS complexes could be crystallized. Although the crystals belonged to the triclinic Laue class and diffracted to only 4.5 Å, the crystallization did provide a second crystal form and also reduced the time between obtaining initial crystals and crystals suitable for data collection. The improvement of crystals of the free form of CitS took more than 1½ years, whereas the complex crystals were refined in less than 3 months. Unfortunately, this process was to no avail, since, neither crystal form could be used to solve the structure. However, the great reduction in time between identification of initial and final crystallization conditions and the numerous additional ankyrin complexes available for testing in the near future could enable structure determination of the first member of the 2-HCT transporter family.

Several other strategies will also be employed to increase the chance of structure determination. Initial experiments with the class of poly oxyethylene ether detergents were promising. The detergent polyoxythylene(8)dodecyl ether ($C_{12}E_8$) promoted a possible different oligomeric assembly in solution. Size exclusion chromatography analysis revealed that the peak was shifted to higher molecular mass and the transporter could be concentrated with a 100-kDa cut-off membrane. This could not be achieved with the detergent DDM, where most of the protein was present in the flow through. This provides an advantage since concentration with a 100-kDa cut-off membrane will allow concentration of the protein-micelle particles but not of empty micelles potentially increasing the reproducibility of the crystallization process. The presence of a higher oligomeric assembly in solution might also result in a different crystal form or, since a tetramer is present in the crystals obtained so far, might improve the quality of those crystals. This has been reported for phage T4 lysozyme, where the synthetic dimerization yielded in numerous novel crystal forms that were not obtainable using the monomeric form of the protein (Banatao *et al.*, 2006, Heinz & Matthews, 1994).

The second strategy is to work with homologous proteins from different species. Similar to the work presented on the ABC transporters, several homologous proteins of CitS will be selected, expressed, purified and crystallized. This strategy is often applied in the field of membrane protein structure determination.

Mutating the desired protein sequence to enhance the crystallization behavior of a membrane protein can be successful, as illustrated with the protein LacY (Abramson *et al.*, 2003). A single site mutation of C154G trapped the transporter in one conformation and enabled the structure determination of a member of the major facilitator superfamily of transporters. A further example where mutations in the protein sequence enhanced the crystallization and diffraction quality is the outer membrane protein OmpA of E.coli. Several mutations had to be introduced in the sequence to produce crystals that were suitable for structure determination (Pautsch & Schulz, 1998). In contrast to LacY, where the mutation in helix V induced a stronger binding interface in the crystals, the residues mutated in OmpA are located in the loop regions of the protein and are not assigned to secondary structural elements. Nevertheless, the mutation of Leu154 to Tyr stabilized a crystallographic contact and led to the formation of crystals with improved diffraction quality. Furthermore, potentially flexible regions in the protein could be stabilized by additional mutations. These sites can be predicted with the entropy server (http://nihserver.mbi.ucla.edu/SER/), but the effect of a specific mutation to the general protein behavior or, more particularly, to the effect on crystallization and diffraction, is, in general, not predictable and all mutations must be tested for expression, folding and crystallization. Nevertheless, homologous proteins or the mutational studies could lead to crystals that are more suitable for determining the structure of the citrate transporter CitS.

Chapter E - Structure of the Fab Fragment f3p4

E.1 Unusual twinning in crystals of the CitS binding antibody Fab fragment f3p4

Daniel Frey, Thomas Huber, Andreas Plückthun and Markus G. Grütter

Unusual twinning in crystals of the CitS binding antibody Fab fragment f3p4

Daniel Frey, Thomas Huber, Andreas Plückthun and Markus G. Grütter*

Institute of Biochemistry University of Zürich, Switzerland.

* E-mail: gruetter@bioc.unizh.ch

PDB-ID: 2v7n

Synopsis:

The structure solution and refinement of pseudomerohedral twinned crystals combined with pseudo NCS of 4 molecules in space group $P2_12_12$ at 1.92Å is reported.

Keywords:

Pseudomerohedral twinning; pseudo non-crystallographic symmetry

Abstract

The structure determination and the refinement at 1.9 Å of the antibody Fab fragment f3p4, selected from the synthetic HuCAL library for binding to the citrate carrier CitS, are described. CDR-L3 carries a trans-proline and thus adopts a conformation between those typical for V_κ and V_λ. The crystals belong to space group $P2_12_12$ and contain 4 molecules in the asymmetric unit with unit-cell parameters of $a = 102.77$, $b = 185.92$ and $c = 102.97$ Å. The self-rotation function revealed a primitive orthorhombic space group with a non-crystallographic symmetry mimicking pseudo-tetragonal symmetry. The particular cell dimension allowed pseudo-merohedral twinning, and the direction of the non-crystallographic symmetry coincides with the observed twinning operator resulting in a 4_12_12 crystal form. Interestingly, the twinning law relates a two-fold screw axis to a two-fold axis.

1. Introduction

The determination of structures of membrane protein structures is still very challenging, as indicated by the yet small number of structures deposited in the PDB database. One bottleneck is the generation of crystals suitable for high resolution structure determination. This is due to the inherent conformational inhomogeneity and flexibility of the proteins, including the loops connecting the transmembrane segments of the detergent-solubilized membrane proteins. One approach to reduce the flexibility is the co-crystallization with antibody fragments (Hunte & Michel, 2002, Ostermeier & Michel, 1997) or other proteins, such as Designed Ankyrin Repeat Proteins (Schweizer *et al.*, 2007, Sennhauser *et al.*, 2006). The antibody fragments or similar proteins provide additional polar surface, which could promote the formation of well ordered crystals.

In all published structures so far, the antibody fragments (scFv or Fab) originate from monoclonal antibodies. The main disadvantage in the generation of monoclonal antibodies by classical hybridoma technology is the inability to control the structural integrity of the membrane protein during the immunization process. The use of phage or ribosome display as selection methods is more advantageous to find binding proteins against the solubilized membrane proteins (Huber *et al.*, 2007, Röthlisberger *et al.*, 2004). Both techniques are faster, compared to eliciting an immune response *in vivo*, but more importantly defined buffer and detergent conditions, which are favorable to the protein of interest, can be used during the selection. In this way, the structural integrity of the membrane protein can be maintained and the likelihood to select binders to conformational epitopes of the native structure of the membrane protein is increased.

From a subset of the human combinatorial antibody library (HuCAL GOLD, Morphosys, Germany), consisting only of the most stable and well expressed frameworks V_H3 and $V_\kappa3$ (Ewert et al., 2003), Fab fragments were selected by phage display to the detergent-solubilized Na^+ citrate symporter CitS of *Klebsiella pneumoniae* (Röthlisberger *et al.*, 2004). CitS is a multitopic membrane protein of the helix-bundle type with relatively short loops connecting the transmembrane segments. The well studied member of the 2-hydroxy-carboxylate transporter (2HCT) family facilitates the symport of citrate and sodium ions, allowing the bacteria to use citrate as sole carbon source. Detergent-solubilized CitS can be functionally reconstituted (Pos and Dimroth, 1996) and therefore it is likely that CitS preserves its natural conformation in the detergent. This allowed us to obtain Fab fragments by phage display that recognize a conformational epitope of the native state of CitS (Röthlisberger *et al.*, 2004).

The Fab fragment f3p4 was selected by phage display and binds specifically to the dimer of CitS with an affinity in the nanomolar range (Huber *et al.*, 2007, Röthlisberger *et al.*, 2004). The co-crystallization of CitS with the Fab fragment f3p4 may not only help to generate crystals suitable for high resolution structure determination, but may also help in phasing the complex by molecular replacement. Thus, the Fab fragment f3p4 was crystallized alone to obtain a high quality molecular replacement search model.

Twinning of macromolecular crystals is not uncommon, and growing data accumulated over the last years show that more and more researchers are facing this problem. Under general, crystals form a continuous assembly of unit cells that are well ordered throughout the crystals. In special circumstances, however, the crystal is made of two or more domains, which are related by symmetry elements, implying higher crystal symmetry; a phenomenon that is called twinning.

Different types of twinning occur in protein crystals. Partial overlapping of diffraction spots is called epitaxial twinning, observed in crystals where the different domains of the crystal are related by a symmetry operator that does not coincide with the unit cell parameters. In contrast, in the case of merohedral twinning, two or more crystal domains are related by an additional symmetry element, creating a perfect superposition of the observed diffraction patterns. Since the individual lattices of each domain are related by additional symmetry, this symmetry element is observed in the combined diffraction pattern. Pseudomerohedral twinning is found in space groups where the unit cell parameters suggest higher symmetry and a different Laue class than actually present in the crystal, e.g., an angle $\beta \approx 90°$ in the monoclinic space group P2 suggests orthorhombic space group P222. The point group of the crystal is 2/m, whereas the lattice has mmm symmetry.

The general problem of pseudo merohedral twinning is that the intensities recorded do not arise from a single crystal but are a summation of intensities of individual domains. In the most common case, where two domains are related by a 180° 2-fold axis rotation, the measured intensities of such a 2-fold related reflections are therefore: $J_{1,obs} = I_1 \cdot (1-\alpha) + \alpha \cdot I_2$ and $J_{2,obs} = \alpha \cdot I_1 + (1-\alpha) \cdot I_2$, where I_1 and I_2 are the un-twinned intensities of individual domains, $J_{1,obs}$ and $J_{2,obs}$ are the observed intensities, and the twinning fraction α is the fractional volume of one domain in the crystal with the second fractional volume corresponding to 1-α. For $\alpha < 0.5$ it is possible to obtain the correct intensities of each reflection and de-twin the data set, but since the error (proportional to $1/(1-2\alpha)$) increases rapidly if α approaches 0.5, de-twinning is not appropriate under these conditions (Yeates & Fam, 1999, Yeates, 1997). Nevertheless, data of twinned crystals with α close to 0.5 and even data of perfectly twinned crystals with $\alpha = 0.5$ can be used to solve a structure by molecular replacement (Yeates & Fam, 1999).

2. Material and methods

2.1. Protein expression and purification

The Fab fragment f3p4 was cloned into the expression vector pMx9 (Rauchenberger *et al.*, 2003). The final construct carried a His_6 tag fused to the C-terminus of the heavy chain. Expression was performed in *E. coli* SB356 (Bass *et al.*, 1996) in 2YT (10 g yeast extract, 16 g tryptone and 5 g NaCl per liter). 30 ml 2YT medium containing 30 µg/ml chloramphenicol and 1.0% glucose were inoculated with a single colony and incubated overnight at 25°C. One liter of dYT medium containing 15 µg/ml chloramphenicol, 50 mM K_2HPO_4 was inoculated with the overnight culture to a starting OD_{600} of 0.1 and incubated in a 5 l shaking flask with baffles at 25°C. Expression was induced at OD_{600} of 0.6 with 1 mM IPTG (final concentration) and continued for 18 hours.

The harvested cells were resuspended in lysis buffer (50 mM Na_2HPO_4 pH 8.0, 150 mM NaCl) containing DNaseI and 1 mg/ml lysozyme and incubated on ice for 30 min. The cells were lysed by passage through a French press and the lysate was cleared by centrifugation (20 min at 40,000*g*) and filtration through a 0.22 µm membrane.

The cleared lysate was applied to a BioCAD system with an IMAC column (Poros 20MC, 7 ml, Applied Biosystems) and an anionic exchange column (Poros 20 HQ, 7 mL, Applied Biosystems) coupled inline. The protein was loaded to the IMAC column and washed with 10 mM Tris pH 8.0, 150 mM NaCl, then high salt (10 mM Tris pH 8.0, 1 M NaCl), then low salt (10 mM Tris pH 8.0, 15 mM NaCl) and a final washing step (10 mM Tris pH 8.0, 15 mM NaCl, 5 mM imidazole). The protein was eluted with 10 mM Tris pH 8.0 and 200 mM imidazole and directly pumped to the anionic exchange column, on which the contaminants were bound to the column, while the Fab fragment f3p4 was recovered in the flow-through. The eluate was further passed over a size-exclusion chromatography column (Superose 6 10/300 GL, GE Healthcare) in 10 mM Tris pH 7.5, 150 mM NaCl and 0.05% (w/v) n-dodecyl-β-D-maltopyranoside to yield a pure and homogeneous protein sample. The protein was concentrated to 12 mg/ml in a centrifugal filter unit (MWCO 30 kDa, Amicon ultra, Millipore) prior to crystallization.

2.2. Crystallization

Crystals of the Fab fragment f3p4 were grown by mixing 1 µl of protein solution with 1 µl of 23% PEG 4000, 50 mM Li_2SO_4, 50 mM Na_2SO_4, 50 mM N-(2-acetamido) iminodiacetic acid in a pH range of 6.3-6.5. The sitting drops were equilibrated by vapor diffusion at 4°C. The crystals appeared overnight and grew to full size within 1 week. Needle like crystals, with a size of 300 µm x 50 µm x 50 µm were flash frozen in liquid propane and 25 % (v/v) ethylene glycol as cryo-protectant.

2.3 Data collection and processing

A data set to 1.9 Å of a single crystal was collected at the X06SA beamline of the Swiss Light Source (SLS) at the Paul Scherrer Institute in Villigen, Switzerland. Diffraction images were indexed and integrated by the program package XDS (Kabsch, 1993). Indexing of a total of 3000 spots with I/s > 5 revealed cell constants of a = 102.77 Å, b = 102.97 Å, c = 185.92 Å and $\alpha = \beta = \gamma = 90°$. Primitive, centered orthorhombic as well as primitive tetragonal were good candidates for the Bravais lattice of the collected crystal. Integration and subsequently scaling values of possible lattices are summarized in Table 1.

Table 1: Scaling statistics

Space group	R_{sym} (%)[1]	R_{meas} (%)[2]	R_{mrgdF} (%)[2]	Completeness (%)	Molecules in the ASU[3]
P1	4.5	6.4	9.0	78.2	16
P2	5.7	6.9	7.9	94.3	8
C2	8.9	11.0	10.3	96.5	8
P222	6.5	7.2	6.0	97.4	4
C222	9.8	10.9	7.8	98.7	4
P4	9.6	10.7	7.6	99.3	4
P422	11.1	11.7	5.7	99.2	2

[1]R_{sym} $\sum|(I_{(h,i)}-I_h| / \sum I_{(h,i)}$; [2]For definition of R_{meas} see (Diederichs & Karplus, 1997) [3]Resulting Matthews coefficient of 2.72 Å^3 Da^{-1} and a solvent content of 54.9%.

Systematic absences were analyzed and indicated the presence of a 2_1 or 4_2 screw axis along the long axis *c*. Additional systematic absences required for the centered orthorhombic space group were not observed. In combination with the self-rotation function (Figure 1) the space group was assumed to be the orthorhombic space group P222$_1$. The additional peak at ω = 90° and ϕ = 45° with an intensity height of 50 % of the origin indicated an imperfect non-crystallographic symmetry axis (NCS) parallel to the diagonal of the *ab* plane.

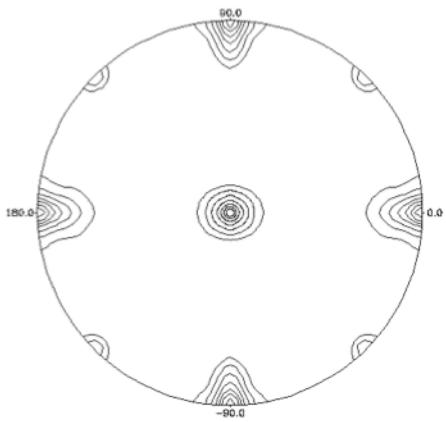

Figure 1: Self-rotation in space group P1

Plot of the κ = 180° section of the self-rotation. The self-rotation function was calculated in space group P1 using the program POLARFN (Collaborative Computational Project Number 4, 1994). The plot was generated with an integration radius of 25 Å and a contour level > 30% of the height of the origin peak using data from 30 to 2 Å. The data reveal P222 symmetry with an additional pseudo two-fold axis at ω = 90 ° and ϕ = 45 °, implying that the additional two fold axis is located parallel to the diagonal of the *ab* plane.

2.4 Structure determination and Refinement

The structure was solved by molecular replacement using Phaser (Collaborative Computational Project Number 4, 1994, Read, 2001) and a search model prepared in the following way: Modeler (Marti-Renom *et al.*, 2000) was used to generate 100 models based on the sequence of the light and the heavy chain of Fab fragment f3p4 and a known Fab fragment structure (PDB-ID: 2fb4, (Kratzin *et al.*, 1989)). To accommodate for the unknown elbow angle between the constant and the variable domain of the Fab fragment, the search model was divided into two parts, the constant and the variable domain. Eight separate domains were used as starting models for molecular replacement, using data from 15 to 3 Å. One correctly assembled Fab fragment was extracted from the first molecular replacement attempt and used as starting model to find the rotation and translation solutions of all four molecules in the asymmetric unit of space group $P222_1$.

Crystal twinning was not considered initially and the structure was refined with CNS (Brünger *et al.*, 1998) using the twinned data without any NCS restraints. After several rounds of model building and refinement the final R-values of $R_{cryst} = 45.3\%$ and $R_{free} = 46.7\ \%$ were unacceptably high. Considering the presence of twinning in the data the molecular replacement trials were repeated in the orthorhombic and tetragonal space groups with all possible permutations. In general, the occurrence of twinning in a crystal should not hinder the structure determination by molecular replacement (Yeates, 1997) and solutions in space groups $P222_1$, $P2_122_1$, $P22_12_1$, $P2_12_12_1$, $P4_1$ and $P4_12_12$ were found (Table 2).

Table 2: Molecular replacement solutions

Space group	LLG[1] / clashes	untwinned R / R_{free}[2]	Twin law and twinned R / R_{free}[3]	
$P222_1$	-468 /5	41 / 43%	-k,h,l	27 / 31%
$P2_122_1$	+1465 /0	34 / 38%	-k,h,l	23 / 26%
$P22_12_1$	-263 /0	38 / 41%	-k,h,l	23 / 27%
$P2_12_12_1$	-1207 /14	41 / 43%	-k,h,l	29 /34%
$P4_1$	-789 /0	40 / 43%	k,h,-l	28 / 32%
$P4_12_12$	-1200 /4	44 / 48%	-	

[1]Log likelihood gain after positional refinement of all molecules. [2]R and R_{free} values reported after refinement with REFMAC without twinning. [3]R and R_{free} values with twinning applied during refinement with SHELX97.

Taking twinning into account the refinement progress was straightforward, and best results were obtained for space group $P2_122_1$. To avoid conflicts during refinement and upon analyzing the structure, the space group was changed to the conventional setting $P2_12_12$. All further refinement steps were done in SHELX97 (Sheldrick & Schneider, 1997) including a two fold twinning around the long axis (h, k, l, → -l, k, h), a starting twin fraction α= 0.37 and local NCS restraints (Uson *et al.*, 1999). The test set was generated in small shells to avoid a correlation between twinned reflections belonging to the test and the reference set (Sheldrick & Schneider, 1997). The refinement statistics are summarized in Table 3.

Table 3: Crystallographic data collection and structure refinement

Data collection	
Space group	$P2_12_12$
Cell dimensions	a = 102.79, b = 185.92, c=102.99
	$\alpha = \beta = \gamma = 90°$
Wavelength (Å)	1.0000
Resolution Range (Å)*	30 - 1.90 (1.92 - 1.90)
Unique reflections (%)*	151549 (3959)
Completeness (%)*	97.4 (83.4)
Rsym (%)*	6.5 (33.1)
Multiplicity*	4.8 (3.3)
I/σ*	17.7 (4.2)
Refinement	
Resolution Range	15 - 1.92
No of reflections / test set	143072 / 4240
Rcryst / Rfree	0.165 / 0.205
$F_O > 4\ \sigma\ (F_O)$	0.159 / 0.199
No. of protein atoms	12905
No. of solvent atoms	350
Twin operator	h,k,l --> -l,k,h
RMS deviation from ideality	
Bonds (Å)	0.006
Angles (Å)	0.022
Ramachandran	
Core (%)	88.7
Allowed (%)	10.6
Generous Allowed (%)	0.4
Disallowed (%)	0.2
Average B values protein	24.2 $Å^2$

*Highest shell values are written in parentheses (1.95 -1.92 Å)

3. Results and Discussion

3.1 Detection of Twinning

The low R-meas during scaling of the data in space group P222 or P4 and the unsatisfactorily high R-values after initial refinement suggested that the Fab fragment f3p4 crystals were twinned.

Twinning can be detected by several methods based on intensity statistics (Dauter, 2003) and in some cases by analyzing the crystal packing (Chandra *et al.*, 1999). In the latter case, the volume of the asymmetric unit in the apparent space group would be too small to contain a complete molecule. Here, however, the tetragonal system with its two point groups 4 and 422 would allow four or two molecules with a Matthews coefficient of 2.72 $Å^3$ Da^{-1} and a solvent content of 54.9% (Matthews, 1968) and, therefore, it was not obvious that the observed tetragonal lattice was a result of twinning.

Closer inspection of the data with the program TRUNCATE of the CCP4 program package (Collaborative Computational Project Number 4, 1994) and the L-test (Padilla & Yeates, 2003) as implemented in the mmtbx.xtriage package of PHENIX (Adams *et al.*, 2004) verified the presence of twinning in the data set. To reduce false positive signals in the calculations the high resolution of the data was cut at 2.2 Å (Lebedev *et al.*, 2006). The cumulative intensity distribution of the measured data of the f3p4 crystals and theoretical statistical data (Rees, 1980) are shown in Figure 2 A. The probability that strong reflections of one lattice superimpose with strong reflections of the second lattice (or weak with weak) is low and, therefore, the overall number of strong and weak reflections is reduced and the cumulative distribution plot adopts a sigmoid curve. The discrepancy in the distribution of centric and acentric reflections is very pronounced and gives a good indication for the presence of twinning. To further prove the presence of twinning in the data set an additional test based on local comparison of the reflection intensities was performed (Padilla & Yeates, 2003). The L-test with $|\mathbf{L}| = (|\mathbf{I_1}-\mathbf{I_2}|)/(\mathbf{I_1}+\mathbf{I_2})$, where I_1 and I_2 are the intensities of reflections 1 and 2 with Miller indices that relate the two reflections close in reciprocal space is, compared to traditional intensity statistics, less sensitive to pseudo-symmetry or anisotropy of the data, both influencing the intensity statistics in a way that twinning detection might fail (Padilla & Yeates, 2003). The calculated mean L-value of 0.37 for the acentric reflections is close to the theoretical value of 0.375 for perfect twinning, compared to an L-value of 0.5 for un-twinned data. Thus, the L-test further established the presence of twinning in the data set.

Several methods have been reported in the literature to give an estimate on the twinning fraction α (Britton, 1972, Lebedev *et al.*, 2006, Padilla & Yeates, 2003, Rees, 1980). In the H-test (Yeates, 1997, Yeates & Fam, 1999) the cumulative fractional difference intensity of twin-related acentric reflections $\mathbf{H} = |\mathbf{I_1}-\mathbf{I_2}|/(\mathbf{I_1}+\mathbf{I_2})$ is plotted against **H** (Figure 2 C). The initial slope of the H-plot indicates that the twinning fraction α is 37%. In contrast, in the Britton plot (Britton, 1972) the fractional negative intensities after de-twinning of the data are plotted against the twinning fraction α. Overestimation of α leads to an increase of the number of negative intensities and α can be extrapolated by a linear fit to this increase. The linear fit of the data with a fractional negative intensity > 0.4 intersects at a twinning fraction $\alpha = 0.36$ with the x-axis (Figure 2 D).

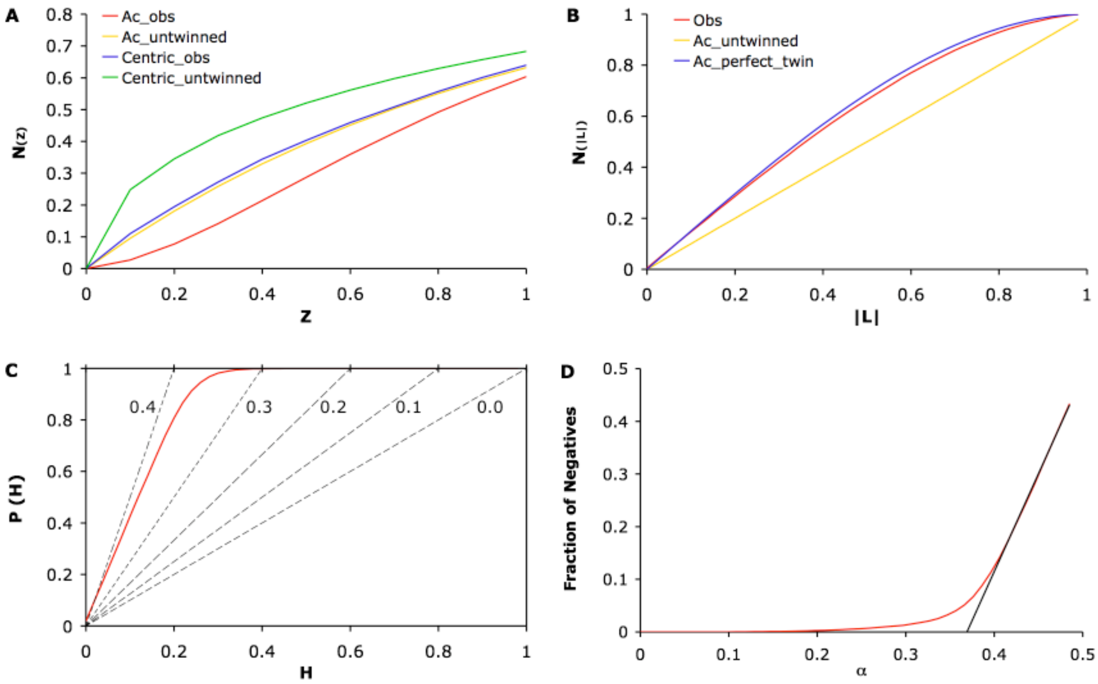

Figure 2: Detection of twinning and determination of α

Detection of twinning based on data set statistics and determination of the twin fraction α. (A) Cumulative intensity distribution as implemented in TRUNCATE. The fraction of average intensity Z is plotted against the fraction of reflections $N_{(Z)}$ with intensities below the average. Both centric and acentric reflections (blue and red) deviate significantly from the theoretical values shown in green and yellow. (B) L-test according to Padilla &Yeates (Padilla & Yeates, 2003). $|L| = (|I_1-I_2|)/(I_1+I_2)$ is plotted against the cumulative distribution of $N_{(|L|)}$ where I_1 and I_2 are intensities of two reflections closely related in reciprocal space. The observed acentric reflections (red) nearly approach the theoretical data of perfect twinned data (blue). The theoretical curve of un-twinned data is indicated in yellow. (C) Determination of the twinning fraction α with the H-test (Yeates, 1997, Yeates & Fam, 1999). $H = (I_1-I_2)/(I_1+I_2)$ is plotted against the cumulative distribution $P_{(H)}$ where I_1 and I_2 are intensities of two reflections related by the twin law. The initial slope of the observed data (red) results in α of 37%. Theoretical lines of different twinning fractions are shown in dashes. (D) Britton plot to determine the twinning fraction α by plotting the fraction of negative intensities after de-twinning against the twinning fraction α (red). The black line shows the fit of the data with α > 0.4 that was used to determine α = 36%.

3.2 Determination of the correct space group

Various space groups appeared reasonable when scaling and merging of the data. Therefore, molecular replacement and the subsequently refinement were used to determine the correct space group. The molecular replacement solutions were initially refined with REFMAC and followed by a twinned refinement in SHELX97. Refinement in space group $P222_1$ including the twin law h, k, l, → -k, h, l did not improve the R-values to satisfactory levels. Space group $P2_122_1$ yielded the lowest R and R_{free} values. Although no systematic absences are present for the screw axis along *a*, it is obvious that the above-mentioned twin law, the high twinning fraction and the two similar cell axes would completely disguise the second screw axis. The presence of the additional tetragonal solution $P4_1$ is explained by the crystal packing of the molecules (Figure 3). A Fab fragment is located between the two molecules that are related by the crystallographic 2_1 screw axis along b. It is rotated by 90° but the translation perpendicular to the screw axis is distorted by ~ 1 Å generating a pseudo 4_1 screw axis along this direction. In combination with the crystallographic 222 symmetry this NCS mimics the 422 symmetry that was observed in the self rotation function (Figure 1). But in contrast to the twinning that generates a $P4_222$ symmetry with a two fold screw axis along *a* or *b*, the 4_1 NCS and the peculiar cell axes mimic the tetragonal space group $P4_12_12$, which was found in molecular replacement but is crystallographically wrong and the reason why the structure could not be refined in this space group.

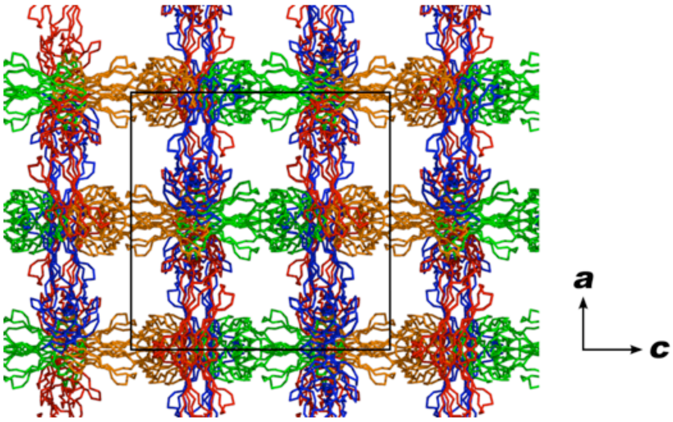

Figure 3 Crystal packing of Fab f3p4

The molecules generate pseudo 4 fold symmetry parallel to *b*. The image shows the projection of the *ac* plane with b normal to the paper plane. Cα traces of the light chains of the four molecules are shown in blue, red, green and orange, and the unit cell is indicated as black square.

Since NCS and the twinning operator coincide, both or just one operator might cause the abnormalities in the intensity distribution of the data. Thus, the RvR plot (Lebedev *et al.*, 2006) based on observed and calculated data was generated after refinement to determine the contributions of the operators. The RvR statistics uses the intensity-based R factor between reflections that are related by a possible twin operator S_{twin}. $R_{twin} = \sum|I_h\text{-}I_{h`}|/\sum (I_h\text{-}I_{h`})$ is calculated for the observed twinned intensities ($R^{obs}{}_{twin}$) and the calculated untwinned intensities ($R^{calc}{}_{twin}$) (Lebedev *et al.*, 2006). $R^{obs}{}_{twin}$ of 0.11 and $R^{calc}{}_{twin}$ of 0.52 indicated indeed a nearly perfect twinning, and furthermore implied that (i) only twinning influenced the data and (ii) the non crystallographic 4_1 axis screw axis along c had no effect on the intensities of the data.

Analysis of deposited PDB data bank entries in space groups susceptible to twinning, reveal that ~30% of a total of 78 twinned structures were pseudomerohedrally twinned (Lebedev *et al.*, 2006), predominantly occurring in space group $P2_1$. However, for space group $P2_12_12$ rare cases of twinning have been reported in the literature. The protein rubredoxin oxygen oxidoreductase crystallized in space group $P2_12_12$ with approximately the same length of *a* and *b*. A twofold NCS along the diagonal of the axes a and b related two molecules in the asymmetric unit, mimicking the space group $P42_12$ (Frazao *et al.*, 1999). A similar twinning was observed in crystals of the β_2 subunit of the Kv1 K^+ channel (Shimamura *et al.*, 2004) with *a* equal *b*. Also there, the diffraction pattern showed an apparent $P42_12$ symmetry caused by crystal twinning along the diagonal in the ab plane. The true $P2_12_12$ space group was determined by molecular replacement with 8 tetramers in P1, where two tetramers formed a face-to-face octamer with 422 symmetry. The additional twofold NCS was also present and observed in the self-rotation function (Shimamura *et al.*, 2004). Although the twinning of the Fab fragment described in this work generates a pseudo-fourfold symmetry around the long axis *c*, emulating with the additional pseudo centering a $P4_122$ symmetry, the twinning is different: Here, the twin operator relates a two fold screw axis with a two fold axis. Therefore, the apparent solution of molecular replacement trials in space group $P4_12_12$ (Table 2) containing two 2_1 screw axes along the short axes *a* and *b*, is false since only one 2_1 screw axis is present along *a* in the true space group. To our knowledge, this unusual twinning has not been reported in the literature so far.

3.3 Structural analysis of the Fab fragment f3p4

The Fab fragment f3p4 crystallized with four molecules in the asymmetric unit. Analysis of the four individual molecules showed identical elbow angles (defined as the angle between the two, not necessarily intersecting, pseudo-dyad axes relating the light (V_L) and heavy (V_H) chain variable domains, and the light (C_L) and heavy (C_H1) chain constant domains (Stanfield *et al.*, 2006), resulting in a rms deviation of 0.5 - 1.1 Å^2 of all C_α atoms. The loop between residues proline 162 and threonine 171 of the constant domain of V_H is the most flexible region in all four Fab molecules. Weak density of this loop was found partly in one of the molecules, whereas the other three were too flexible to be modeled.

The overall structure of the Fab fragment f3p4, which was selected from the HuCAL GOLD library is very close to the proposed models (PDB-ID 1DH5 and 1DHU) of the consensus V_H3 and V_k3 frameworks (Knappik *et al.*, 2000). The rms deviation of all main chain atoms is 1.31 Å (min. 0.03, max. 6.55) for V_H3 and 0.81 Å (min. 0.08, max. 2.83) for V_k3.

The structure of the Fab fragment f3p4 is the first published crystal structure of a Fab fragment originating from the synthetic HuCAL library. It supports the design of the library to contain antibodies corresponding to the natural repertoire. The deviations from the modeled structures (which had been built based on natural antibodies) are minor and limited to loop regions (L47-51, L83-88, L99-102 and H47-50, Honegger numbering, Honegger & Plückthun, 2001) outside of the central β-sheets, in the side chain orientation of solvent accessible charged residues and in the complementary determining regions CDR-L1, CDR-L3, CDR-H1 and CDR-H3. These deviations are within the range of structural variations of antibody structures obtained by hybridoma technology, i.e., are also found in the natural repertoire.

The overall conformation of the CDR-L1 corresponds well to the typical hairpin structure observed in human V_K3. The upper loop of this CDR has a high conformational variability in natural antibodies and the detected conformation of this loop is well within the range of human antibody structures (See, e.g. PDB-IDs: 1BWW, 1DNO, 1HEZ or 1JV5) as well as of the V_K3 model. The CDR-L1 is also involved in crystal contacts. CDR-L2 hardly varies between antibody V_L domains.

The typical omega shape of the V_K CDR-L3 is determined by a cis-proline at position L136 (95, Kabat numbering), which is carried by more than 90% of the germline V_K sequences and about 60% of rearranged human V_K antibody sequences (A. Honegger, unpublished results), as it is the last residue encoded by most of the kappa V-segments. In the Fab fragment f3p4, selected from the HuCAL V_K3 V_H3 subset (Röthlisberger *et al.*, 2004), there is a serine at this position. Therefore, the CDR-L3 does not adopt an omega loop conformation. At position L137 (96, Kabat numbering), a trans-proline is present and stabilized by a π-π interaction with the side chain of tryptophan H54 of the heavy chain. Residue L137 would correspond to the first residue of the kappa J1 segment, and would encode a Trp residue (or another aromatic or hydrophobic in all J segments), but this residue is frequently mutated in V-J recombination. In contrast to V_K domains, V_λ normally do not have a proline residues at this position in CDR-L3, giving the loop a more hair-pin like shape. The presence of this trans-proline in f3p4 prevents a fully V_λ CDR-L3 like hairpin conformation, giving it an intermediate shape (Figure 4).

This amino acid sequence of f3p4 is as expected from the design, as in the HuCAL library the frequency of serine is 20 % at position L136 (80 % of proline) and position L137 is fully randomized (except for cysteines). Even though this amino acid composition is not predominant in natural antibodies, it can be found, e.g. in the antibody PDB entry 1C1E, which has been selected from hybridomas (Figure 4).

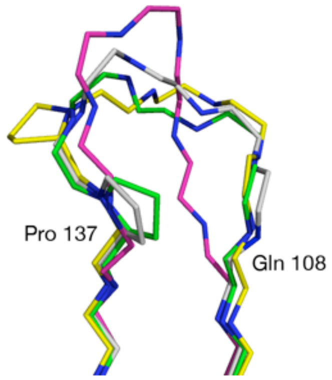

Figure 4: Structure of CDR-L3

The superposition of the main chain atoms (For clarity, carbonyl atoms are not shown) of CDR-L3 of the fab f3p4 (grey) with typical members of V_K and V_λ is shown. V_λ members adopt a hairpin like structure (magneta, PDB entry 8FAB). V_K contains predominantly a proline at position L136 forcing an omega loop conformation (yellow, PDB entry 1AY1). The conformation of the CDR-L3 of f3p4 is an intermediate, but is present in natural occurring antibodies (green, PDB entry 1C1E) with a proline residue at L136. The molecules contain equal numbers of residues in CDR-L3 and Cα atoms of Cys 106 – Gly140 (Honegger numbering) were used for the superimposition.

In the V_H3 heavy chain of f3p4, CDR-H1 shows a conformation typical for this subclass. The glycine at position H65 (CDR-H2, Honegger numbering) is the center of a short hairpin loop ranging from amino acids H59-67. This CDR shows the highest mobility in the structure in all four molecules, reflected by two major conformations and higher B-values (Figure 5).

The CDR-H3 is by far the most diverse in sequence variability and length. Its conformation is therefore difficult to predict. The hydrogen bonding between arginine H108 and aspartate H137, giving rise to the typical kink of the CDR-H3, is also observed in the structure of f3p4. A relatively open conformation of this CDR is found, allowing residues of the target to penetrate into the groove at the VL/VH interface (Figure 6). The CDR-H3 is also involved in crystal contacts.

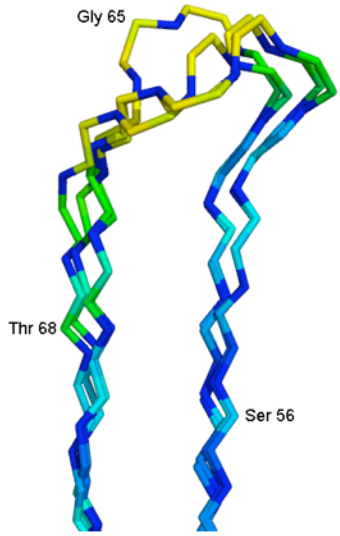

Figure 5: Variability in CDR-H2

Superimposition of the 4 individual CDR-H2 segments of the Fab f3p4 structure. The hairpin loop around Gly H65 adopts different conformations in the four subunits and high B-factors are found in this region. Main chain trace without the carbonyl atoms is shown. Atoms are colored in light blue, green and yellow according to low, medium and high mobility, respectively. N atoms are colored in dark blue.

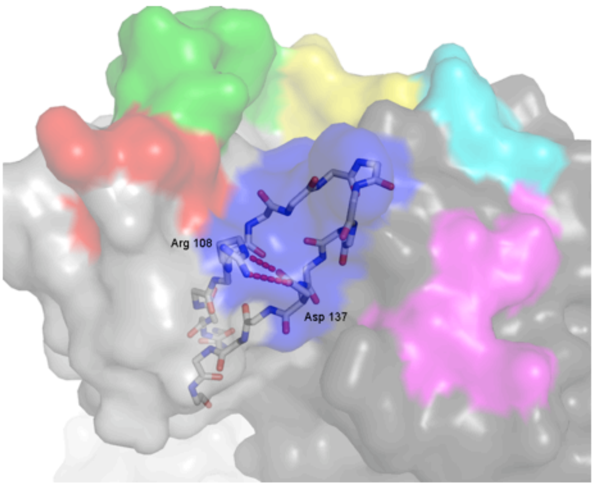

Figure 6: CDR-H3

Surface representation of the binding region. CDRs H1, H2, H3, L1, L2 and L3 are colored in red, green, blue, cyan magenta and yellow, respectively. CDR H3 is shown in stick representation and the conserved hydrogen bonds between Ar108 and Asp137 are shown as red dashes.

We also discuss two side chain conformations: Phe L101 is rather solvent exposed in the modeled structure of V_K3 but buried in the f3p4 structure. Both conformations occur in the antibody structures, and therefore its conformation cannot be predicted. Arg H82 was previously modeled to be solvent exposed in V_H3 (PDB-ID: 1DH5). In the f3p4 structure, this residue is fully buried and forms hydrogen bonds to the side chain of Asn H84 and to the main chain of Ser H39. The buried conformation and the hydrogen bonding pattern is observed in the majority of antibody V_H structures.

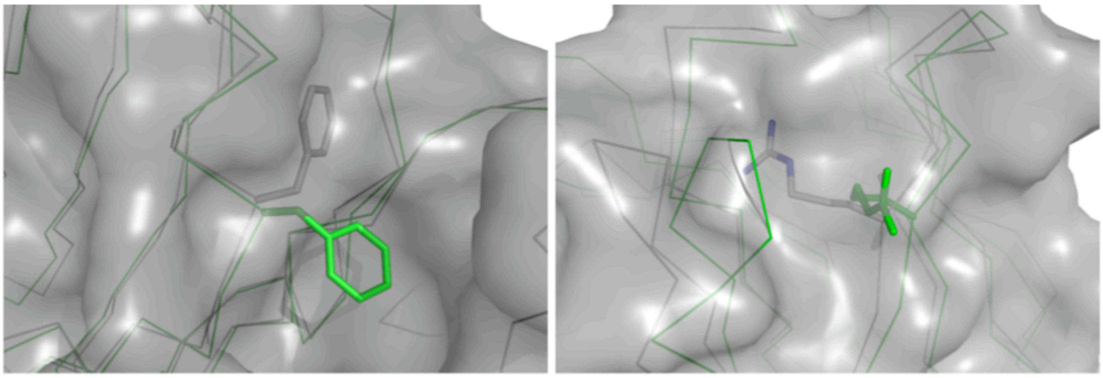

Figure 7: Solvent accessibility of PheL75 and ArgH82.

Left image: PheL75 is buried in the Fab f3p4 structure (grey) and solvent accessible in the model (green, 1DHU). Right image: Residue ArgH75 is buried in the Fab f3p4 structure (grey) and more solvent exposed in the model (green, 1DH5). Surface was calculated using the coordinates of the Fab f3p4 structure. Models 1DH5 and 1DHU were superimposed with one molecule of the Fab f3p4 structure using the program COOT (Emsley & Cowtan, 2004, Krissinel & Henrick, 2004).

The Fab fragment f3p4 precipitates at concentrations above 13 mg/ml in buffers containing 150 mM or more NaCl. This precipitate can be re-dissolved by simply dialyzing away the salt and thus appears to be an association of native molecules. We have selected four Fab fragments (Röthlisberger *et al.*, 2004) binding to CitS and only two of them show this effect. The framework of all four Fab fragments is the same (V_K3/V_H3) and they differ only in the CDRs. CDR-L1 (residues L30, L32, L33, L39 and L40), CDR-L2 (L57, L58, L68 and L69), together with an exposed tryptophan

(L133) of CDR-H3, are involved in crystal contacts to the constant domain of the light chain (residues L190, L192, L193, L228-232, L234, L236, L244, L245, L247, L249 L251 and L252). This contact consists to 50% of non-polar residues, four hydrogen bonds and covers an area of about 450 $Å^2$. By comparing the CDRs of all four selected Fab fragments, we found that both Fab fragments, which do not show precipitation, have an CDR-L1 that is shortened by one amino acid and missing the tyrosine at position L40. CDR-L1 might therefore not only contribute to the crystal contact but also be responsible for the precipitation at high concentration in the presence of salt.

4. Concluding remarks

This work represents the first experimental structure solution of a Fab fragment derived of the fully synthetic HuCAL GOLD® antibody library, and may thus help in further improving antibody modeling. The crystals of this particular Fab fragment exhibit a pseudomerohedral twinning combined with a pseudo non-crystallographic symmetry. The twinning operator and the NCS coincide, a phenomenon that was reported before (Barends & Dijkstra, 2003, Cao *et al.*, 2006, Golinelli-Pimpaneau, 2005, Larsen *et al.*, 2002, Rudolph *et al.*, 2004), although the correlation of a two-fold axis and a two-fold screw axis, as observed here, was to our knowledge not reported before. In general, the mentioned twinning detection tools and their implementation in the standard crystallographic software allow all users to detect a possible twinning even in those space groups where twinning is not expected.

Acknowledgement

We gratefully acknowledge the Swiss Light Source, Paul Scherrer Institute, Villigen, Switzerland, for providing synchrotron beamtime, Peer Mittl for helpful discussions and Stefan Dudli and Dominik Hänni for help in protein purification and crystallization.

References

Adams, P. D., Gopal, K., Grosse-Kunstleve, R. W., Hung, L. W., Ioerger, T. R., McCoy, A. J., Moriarty, N. W., Pai, R. K., Read, R. J., Romo, T. D., Sacchettini, J. C., Sauter, N. K., Storoni, L. C. & Terwilliger, T. C. (2004). Journal of Synchrotron Radiation 11, 53-55.

Barends, T. R. & Dijkstra, B. W. (2003). Acta Crystallographica Section D-Biological Crystallography 59, 2237-2241.

Bass, S., Gu, Q. & Christen, A. (1996). J Bacteriol 178, 1154-1161.

Britton, D. (1972). Acta Crystallographica Section A 28, 296-297.

Brünger, A. T., Adams, P. D., Clore, G. M., DeLano, W. L., Gros, P., Grosse-Kunstleve, R. W., Jiang, J. S., Kuszewski, J., Nilges, M., Pannu, N. S., Read, R. J., Rice, L. M., Simonson, T. & Warren, G. L. (1998). Acta Crystallographica Section D-Biological Crystallography 54, 905-921.

Cao, P., Gong, Y., Tang, L., Leung, Y. C. & Jiang, T. (2006). Journal of Structural Biology 154, 327-332.

Chandra, N., Acharya, K. R. & Moody, P. C. (1999). Acta Crystallographica Section D-Biological Crystallography 55, 1750-1758.

Collaborative Computational Project Number 4 (1994). Acta Crystallogr D Biol Crystallogr. 50, 760-763.

Dauter, Z. (2003). Acta Crystallographica Section D-Biological Crystallography 59, 2004-2016.

Diederichs, K. & Karplus, P. A. (1997). Nature Structural Biology 4, 269-275.

Emsley, P. & Cowtan, K. (2004). Acta Crystallographica Section D-Biological Crystallography 60, 2126-2132.

Frazao, C., Sieker, L., Coelho, R., Morais, J., Pacheco, I., Chen, L., LeGall, J., Dauter, Z., Wilson, K. & Carrondo, M. A. (1999). Acta Crystallographica Section D-Biological Crystallography 55, 1465-1467.

Golinelli-Pimpaneau, B. (2005). Acta Crystallographica Section D-Biological Crystallography 61, 472-476.

Huber, T., Steiner, D., Röthlisberger, D. & Plückthun, A. (2007). J Struct Biol.

Hunte, C. & Michel, H. (2002). Current Opinion in Structural Biology 12, 503-508.

Kabsch, W. (1993). Journal of Applied Crystallography 26, 795-800.

Knappik, A., Ge, L., Honegger, A., Pack, P., Fischer, M., Wellnhofer, G., Hoess, A., Wölle, J., Plückthun, A. & Virnekäs, B. (2000). Journal of Molecular Biology 296, 57-86.

Kratzin, H. D., Palm, W., Stangel, M., Schmidt, W. E., Friedrich, J. & Hilschmann, N. (1989). Biological Chemistry Hoppe-Seyler 370, 263-272.

Krissinel, E. & Henrick, K. (2004). Acta crystallographica 60, 2256-2268.

Larsen, N. A., Heine, A., de Prada, P., Redwan el, R., Yeates, T. O., Landry, D. W. & Wilson, I. A. (2002). Acta Crystallographica Section D-Biological Crystallography 58, 2055-2059.

Lebedev, A. A., Vagin, A. A. & Murshudov, G. N. (2006). Acta Crystallographica Section D-Biological Crystallography 62, 83-95.

Marti-Renom, M. A., Stuart, A. C., Fiser, A., Sanchez, R., Melo, F. & Sali, A. (2000). Annual Review of Biophysics & Biomolecular Structure 29, 291-325.

Matthews, B. W. (1968). Journal of Molecular Biology 33, 491-497.

Ostermeier, C. & Michel, H. (1997). Current Opinion in Structural Biology 7, 697-701.

Padilla, J. E. & Yeates, T. O. (2003). Acta Crystallographica Section D-Biological Crystallography 59, 1124-1130.

Rauchenberger, R., Borges, E., Thomassen-Wolf, E., Rom, E., Adar, R., Yaniv, Y., Malka, M., Chumakov, I., Kotzer, S., Resnitzky, D., Knappik, A., Reiffert, S., Prassler, J., Jury, K., Waldherr, D., Bauer, S., Kretzschmar, T., Yayon, A. & Rothe, C. (2003). J Biol Chem 278, 38194-38205.

Read, R. J. (2001). Acta Crystallographica Section D-Biological Crystallography 57, 1373-1382.

Rees, D. C. (1980). Acta Crystallographica Section A 36, 578-581.

Röthlisberger, D., Pos, K. M. & Plückthun, A. (2004). FEBS Letters 564, 340-348.

Rudolph, M. G., Wingren, C., Crowley, M. P., Chien, Y. H. & Wilson, I. A. (2004). Acta Crystallographica Section D-Biological Crystallography 60, 656-664.

Schweizer, A., Roschitzki-Voser, H., Amstutz, P., Briand, C., Gulotti-Georgieva, M., Prenosil, E., Binz, H. K., Capitani, G., Baici, A., Plückthun, A. & Grütter, M. G. (2007). Structure 15, 625-636.

Sennhauser, G., Amstutz, P., Briand, C., Storchenegger, O. & Grütter, M. G. (2006). PLoS Biol 5, e7.

Sheldrick, G. M. & Schneider, T. R. (1997). Methods in Enzymology 277, 319-343.

Shimamura, T., Shamotienko, O., Akhtar, S., Dolly, J. O. & Iwata, S. (2004). Acta Crystallographica Section D-Biological Crystallography 60, 912-914.

Stanfield, R. L., Zemla, A., Wilson, I. A. & Rupp, B. (2006). J Mol Biol 357, 1566-1574.

Uson, I., Pohl, E., Schneider, T. R., Dauter, Z., Schmidt, A., Fritz, H. J. & Sheldrick, G. M. (1999). Acta Crystallographica Section D-Biological Crystallography 55, 1158-1167.

Yeates, T. O. (1997). Methods in Enzymology 276, 344-358.

Yeates, T. O. & Fam, B. C. (1999). Structure 7, R25-29.

Chapter F - Appendices

F.1 Crystallographic Methods

Data Collection

Collection of data at home or at the synchrotron was always carried out in a similar manner: a couple of images were first taken to determine a suitable oscillation angle, exposure time and distance between crystal and detector before collecting a full data set from the crystal.

Protein crystals contain an inherent amount of disorder which limits diffraction quality. Data processing of the resulting spots on the diffraction images requires that each spot can be separated from its neighbors. It is not possible to correctly measure the intensity of spots that overlap with neighboring reflections on consecutive images. The oscillation angle selected for data collection must be chosen to reduce the number of overlapping spots. The separation of the spots on individual images can be achieved by varying the detector to crystal distance. An increase in distance generates greater spot separation but decreases the maximum resolution obtained. A good balance between spot separation and resolution must therefore be found.

The high intensity of synchrotron radiation generally allows diffraction to higher resolution, but it also increases the radiation damage of the crystal. Several kinds of radiation damage are known, including reduction of disulfide bonds, decarboxylation or radical formation (Burmeister, 2000, Garman, 2003, Ravelli & McSweeney, 2000). Reducing the beam intensity, decreasing the time of exposure, or translation of the crystal after several frames can be employed to counteract the radiation problem and ensure a continuous data quality during the measurement.

Data Processing

Data were processed using one of two programs; XDS (Kabsch, 1993) or the HKL suite (Otwinoswki & Minor, 1997). To determine the correct unit cell and crystal orientation, parameters such as the wavelength of the radiation, the angle over which the crystal was oscillated per frame, the direct beam position, detector characteristics and detector to crystal distance must be provided.

Processing using the Program Suite HKL

Strong reflections from a single image were used to determine the crystal orientation and the correct unit cell. The resulting predicted pattern was compared to the diffraction pattern of the indexed frame. The following variables were adjusted before processing the first frame: direct beam position, detector to crystal distance, mosaicity, spot and box sizes. The parameters obtained from the first image were used to process the remaining images.

The scaling and merging of the data, as well as the global refinement of the crystal parameters, were performed by Scalepack (Otwinoswki & Minor, 1997). In the first run, the program rejects spots from the images and stores them in a separate file. After adjusting the error model, the program was rerun several times excluding the rejected spots. Once no more reflections were rejected and the goodness of fit (Chi^2) was around 1, the scaling was stopped. The R_{sym} and the signal to noise ratio I/□ indicate the quality of the data. R_{sym} compares intensities of reflections that have been measured more than once (multiplicity) or of their symmetry equivalents.

$$R_{sym}(I) = \frac{\sum_{hkl}\sum_{i}\left|I_i(hkl) - \bar{I}(hkl)\right|}{\sum_{hkl}\sum_{i}\left|I_i(hkl)\right|}$$

$I_i(hkl)$ defines the observed intensity of a measured reflection and $\bar{I}(hkl)$ the mean intensity of the observation i. A low R_{sym} value indicates that there is a good agreement between symmetry related reflections, whereas high R_{sym} values occur if the data was incorrectly indexed or the measurement of the intensities was inaccurate. A high multiplicity with a low R_{sym} value indicates that the reflection intensities have been reliably measured. Other indicators of data quality are the signal to noise ratio I/□ and the completeness of the data. In general data were truncated at I/□< 2, R_{sym} >50% or if the completeness dropped below 80%.

Data Processing with XDS

The program package XDS (Kabsch, 1993) refines the crystal and experimental parameters and integrates the intensities (subprogram XDS), calculates statistics for the data (subprogram XSCALE) and converts the output file to formats suitable for crystallographic refinement programs (subprogram XDSCONV). The parameters used for data collection were input to XDS. Initially strong reflections from the first 5 to 10° were analyzed to determine the space group and cell parameters. These cell parameters are then provided for subsequent examination of the entire data set. The resulting reflection file is merged and scaled with XSCALE and the data is evaluated according to the values mentioned in the previous section (Processing using the program suite HKL).

Structure Determination

Calculation of an electron density map requires structure factors F_{hkl}. F_{hkl} is a periodic function, and therefore can be described by its amplitude, frequency and phase. The frequency corresponds to the frequency of the X-ray source and the amplitude is proportional to the square root of the measured intensity, I_{hkl}, of the reflection. The phase angle can not be directly measured in a single experiment. The phasing information can be obtained in several ways, often requiring the presence of additional heavy atoms in the unit cell. Single or multiple isomorphous replacement (SIR or MIR) use data that have been colleted in the presence and absence of heavy metals. Combining the native and derivatized data sets, the heavy metals can be placed in the unit cell and their positions are used to calculate initial phases. The methods single- and multi- wavelength anomalous dispersion (SAD, MAD) take advantage of the signal from anomalous scatterers at particular wavelengths. These techniques depend on a small difference in anomalous signal that leads to disruption of Friedel`s law. In MAD two or more wavelengths are collected and phases can be obtained by measuring differences in the intensities due to the changes in anomalous behavior at the different wavelengths.

A further method to obtain phase information can be used if a similar protein structure is available and identified. This method, called molecular replacement technique is explained in detail in the next chapter.

The structure factor F_{hkl} is the sum of all scattering atoms in the unit cell of a crystal.

$$\mathrm{F_{hkl}} = \sum_{N} \mathrm{f_i} \exp\left[2\pi \mathrm{i}\left(\mathrm{hx_i + ky_i + lz_i}\right)\right] \exp\left[-\mathrm{B_i} \sin^2\theta / \lambda^2\right]$$

The atomic scattering factor f_i of each atom depends on the number of electrons of an atom. The last term in the equation, the B or temperature factor, represents the thermal and statistic disorder of the crystal lattice. f_i and the B-factor decrease with higher resolution and thereby limit the intensity of the structure factor F_{hkl}. The electron density map can be calculated by performing a Fourier transform of the individual structure factors:

$$\rho(x,y,z) = \frac{1}{\mathrm{V}} \sum_{h} \sum_{k} \sum_{l} \mathbf{F}(\mathrm{hkl}) \exp(-2\pi \mathrm{i}(\mathrm{hx + ky + lz}))$$

where V is the volume of the unit cell and ρ the corresponding electron density at the grid point x, y and z. **F** (hkl), the structure factor of the reflection (hkl), is the product of the amplitudes |F| obtained from the measured intensities **I**(hkl) where **I** equals $|F|^2$ and the corresponding phase angle α (**F**(hkl)=|F|exp[iα]). α is determined by molecular replacement, single or multiple isomorphous replacement or anomalous dispersion methods (SAD, MAD).

Structure Determination by Molecular Replacement

A method to overcome the “phase problem“ is: retrieve the phases from the structure factors of a known protein structure and use these phases as initial estimates for the new structure to be determined. This method is called molecular replacement and as the number of structures deposited in the protein data bank is growing rapidly, it is becoming more and more important.

To apply the phases of the known structure to the unknown molecule, both models have to be superimposed. Therefore the known structure (trial model) has to be positioned in the unit cell of the unknown structure using a rotation and translation function. The superposition is carried out using the Patterson function $P(\vec{u})$, which is a Fourier summation of intensities with all phase angles equal to zero. Patterson functions are calculated for both the known and unknown molecule. The obtained Patterson map contains two types of vectors: self vectors, which are vectors between atoms in an individual molecule; and cross vectors, which are vectors between atoms in one molecule in the crystal and atoms in an other molecule. Self vectors are used to define the correct rotational orientation of the trial model whereas cross vectors are used to translate the previously rotated trial model in the unit cell of the unknown protein. A schematic representation is show in Figure 72.

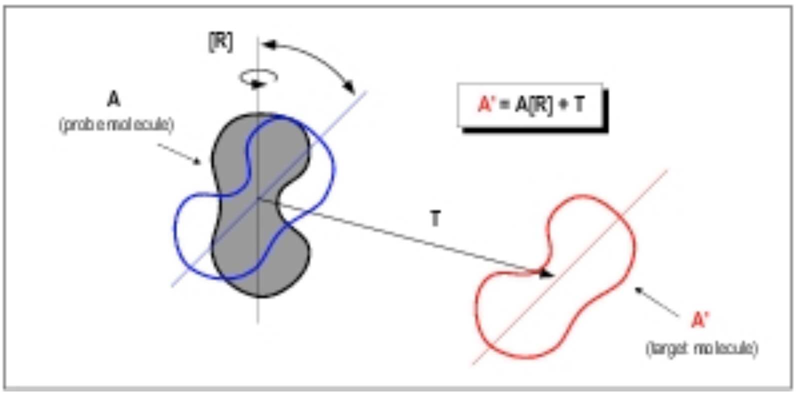

Figure 72: Molecular replacement. A graphic representation of the main principle of the molecular replacement method is shown. The phasing molecule (probe) must be rotated and translated to fit the unknown molecule (target molecule) in the asymmetric unit of the crystal. R is the rotation matrix and T the translation vector.
(http://daffy.bioc.cam.ac.uk/~dima/whitepapers/mr-in-action/node4.html)

The software package AMoRe (Navaza, 1994) was applied to solve the structure of DDAH-1. 1H70, a close homologue of DDAH-1, was used as the search model (Murray-Rust *et al.*, 2001). The cross-rotation function, defined as the overlap of the Patterson functions of the unknown structure and the phasing model, was calculated using a sphere radius for the model of 25 Å and a resolution range of 10.0-4.0 Å. The search model had previously been reduced to a poly serine model with the program MOLEMAN (G.J. Kleywegt; 1992-2004, Uppsala University, Uppsala, Sweden, unpublished program). The poly serine model is useful since it provides more information than a poly alanine would. The best 99 solutions from the rotational search were then applied to the translational search function. The top solutions of the translational search were minimized using the rigid body refinement function implemented in the AMoRe program package.

Model Building and Refinement

Refinement of a crystal structure is an iterative process of adjusting the model to find a better agreement between the calculated and the observed structure factors. The quality of the model depends upon the target function used for the refinement. The two target functions, least-square (LSQ) and maximum likelihood (ML), are used in macromolecular structure determination. The programs minimize a residual value E, which is the weighted sum of squared deviations between the observed $(|F_o|)$ and the calculated $(|F_c|)$ structure-factor amplitudes, including a relative scaling factor k and w_a, which is a weight between experimental and geometric restraints:

$$E^{LSQ} = E_{restraints} + w_a \sum_{hkl} \left(|F_o| - k|F_c|\right)^2 .$$

$E_{restraints}$ are geometric (bond length, bond angle, planarity and atomic repulsion) restraints included in the refinement. Although the atomic model is improved, problems arise because such a least square residual is poorly justified, especially in the early stage of refinement. The value that is minimized is the root mean square deviation between the model electron density and the electron density computed from Fourier coefficients

$|Fo|\exp(i\alpha_c)$, where the phases α_c are obtained from the model and $|Fo|$ are derived from the measured intensities. This deviation can be minimized either by improving the model or by introducing systematic errors that erase differences in the electron density. Due to the unfavorable parameter-to-observation ratio, the data are typically overfitted, which means that systematic errors are introduced (Pannu & Read, 1996).

A more general approach, which also applies to incomplete or error containing models, uses maximum likelihood (ML) functions. The principle of ML is that the quality of the model is judged by its consistency with the measured observations. Consistency means that a correct model would reproduce the observed data points with a high probability. The effects of model errors (misplaced or missing atoms) on the calculated structure factors are quantified with σ_a values, which correspond roughly to the fraction of each structure factor that is expected to be correct. A random test set of reflections (5-10%), which is excluded from the refinement process, is used to calculate σ_a. These cross-validated values are used to compute $\langle|Fo|\rangle^{cv}$ and the corresponding variance ($\sigma^2_{ML}{}^{cv}$). The target function that is minimized is given by:

$$E^{ML} = E_{restraints} + w_a \sum_{hkl \in workingset} \left(\frac{1}{\sigma^2_{ML}{}^{cv}}\right)\left(|Fo| - \langle|Fo|\rangle^{cv}\right)^2 .$$

The advantages of ML are that the false minima are less pronounced, the radius of convergence is larger and overfitting compared to LSQ refinements is reduced (Adams *et al.*, 1997).

The fit of the calculated and experimentally observed structure factors is described by the R-factor:

$$R = \frac{\sum_{hkl} \left||F_{obs}| - k|F_{calc}|\right|}{\sum_{hkl} |F_{obs}|}$$

The temperature Factor (B-factor)

The temperature or B-factor represents the thermal and statistic disorder of the crystal. The intensity of a reflection is reduced according to:

$$I = I^{o} \exp [-B (\sin^2\theta / \lambda^2)]$$

where B equals $8\pi^2u^2$. u is the root-mean-square displacement of the atom around the ideal position at 0 Kelvin. This factor only reduces the intensity of the observation and does not change the sharpness or shape of the spot in the diffraction experiment. The overall B-factor can be estimated in early stages during structure determination. The intensities determined from diffraction images are divided into a series of shells according to the resolution (d = $\lambda/2\sin\theta$) and the mean intensity $\bar{I}$ is calculated in each shell. A graph ln ($\bar{I}/\bar{f}^2$) where f are the individual atomic scattering factors against $(\sin\tilde{\theta}\lambda)^2$ assuming a random distribution of the atoms results ideally in a straight line with a slope of –B and is called a Wilson plot (Wilson, 1949). The Wilson plot for a real protein structure at a resolution higher than 3 Å resolution deviates from the theoretical straight line because of secondary structural elements, which disturb the statistical intensity distribution due to the regular arrangements of atoms.

In refinement, the positions and the B factor of each individual atom are determined. However, the low observation to parameter ratio only allows the determination of B of each atom for resolutions better than ~ 2.5 Å. At lower resolution the B factor is therefore calculated either separately for side and main chains or only one B is calculated for each amino acid. At high resolution the motions of the atoms can be described in more detail. Here, the distortion of the atomic electron density might not be spherical (isotropic), but could vary depending on the direction. Therefore the anisotropic B-factor is introduced, which is displayed as an ellipsoid around the central atom position. In total, the number of parameters to be defined therefore increases to 9; three parameters locate the individual atom in space (x, y, z) and six parameters define the ellipsoid of the anisotropic B-factor U_{11}, U_{22}, U_{33}, U_{12}, U_{13} and U_{23}.

F.2 Biochemical Methods

Protein and DNA Markers

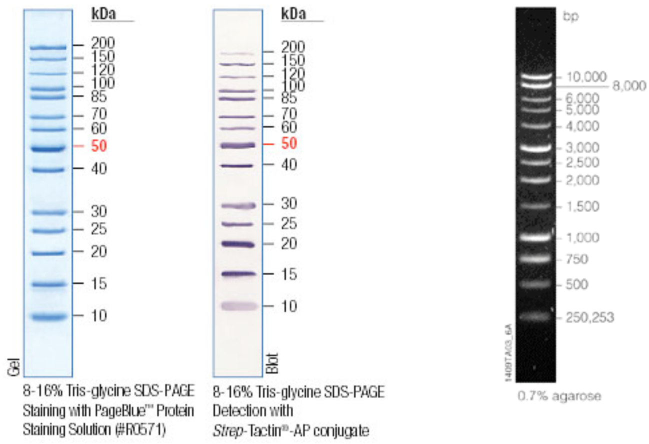

Figure 73: Protein and DNA Marker used in this work. Protein ladder was purchased from Fermentas and 1kb DNA ladder was obtained from Promega.

ATP Hydrolysis Assay

1-12 µl of the sample was inoculated with 100 µl ATPase buffer (50 mM Hepes, 2 mM EGTA, 2 mM DTT, 5 mM NaN_3, 1 mM Ouabain, $MgCl_2$, pH 7.4) and ATP was added to a final concentration of 5 mM. Care had to be taken that the ATP solution was buffered to ~pH 6-7 with NaOH. The hydrolysis of ATP was stoped after 20 min at 37°C by adding 112 µl of a 5% SDS solution. The solution was diluted by adding 1 ml water and 400 µl ammonium molybdate solution (1.25% NH_4) MoO_4 in 1.25 M H_2SO_4). After vigorous mixing the color development was started by adding a 1% solution of the Fiske-Subbarow reducer. After 10 min the samples were measured at 660 nm. Linearity of the activity assay was checked with a 10 mM phosphate solution. In the case of mBsep, the ATPase activity was enlarged by the addition of the substrate taurochendeoxycholate in a final concentration of 30 – 50 µM.

F.3 ABC Transporters

mBsep - Constructs

mBsep pFASTBac HTb: full cDNA of mBsep cloned in pFASTBac HTb (Ehe I/ Not I). Expression vector used to infect insect Sf9 cells with baculovirus.
The N-terminal extension of MSYYHHHHHHDYDIPTTENLYFQG contained the His-tag as well as the protease cleavage site for the TEV protease. The vector was obtained from B. Stieger of the University of Zurich.

mBsep pET20NHT: full length cDNA of mBsep in the bacterial expression vector pET20 NHT.
Sense primer:
5`-GCTGAATTCATGTCTGACTCAGTGATTCTTCGCAGTG-3`
Antisense primer:
5`-GCTATCAGCGGCCGCTTCAACTGATGGGGGCTCCAG-3`

mBsep pET28NH: full length cDNA of mBsep in bacterial expression vector pET28 NH.
Sense primer:
5`-GCTGAATTCATGTCTGACTCAGTGATTCTTCGCAGTG-3`
Antisense primer:
5`-GCTATCAGCGGCCGCTTCAACTGATGGGGGCTCCAG-3`

mBsep pET28ch: full length cDNA of mBsep in bacterial expression vector pET28 ch.
Sense primer:
5`-GCTGAATTCATGTCTGACTCAGTGATTCTTCGCAGTG-3`
Antisense primer:
5`-GCTATCAGCGGCCGCTTCAACTGATGGGGGCTCCAG-3`

mBsep I pET 20ch: amino acids 1-755 of mBsep in pET20ch vector.
Sense primer:
5`-GCTGAATTCATGTCTGACTCAGTGATTCTTCGCAGTG-3`
Antisense primer:
5`-GCTACTGCGGCCGCTTATGTAGGGCCATTCTGAG-3`

mBsep I pET20NHT: amino acids 1-755 of mBsep in pET20 NHT vector.
Sense primer:
5`-GCTGAATTCATGTCTGACTCAGTGATTCTTCGCAGTG-3`
Antisense primer:
5`-GCTACTGCGGCCGCTTATGTAGGGCCATTCTGAG-3`

mBsep II pET 20ch: amino acids 756-1320 of mBsep in pET20ch vector.
Sense primer:
5`-GCTGAATTCCTGGTAGGGGCTTTGTGTGC-3`
Antisense primer:
5`-GCTACCTCGAGACTGATGGGGGCTCCAGTGATG-3`

mBsep II pET20NHT: amino acids 756-1320 of mBsep in pET20 NHT vector.
Sense primer:
5`-GCTGAATTCCTGGTAGGGGCTTTGTGTGC-3`
Antisense primer:
5`-GCTACCTCGAGACTGATGGGGGCTCCAGTGATG-3`

pEAK8 mBsep: full length mBsep in the vector pEAK8 designed for expression in mammalian cells. The construct contained an additional N-terminal 6xHis-tag plus TEV recognition site.
Sense primer:
5`-GCTACAAGCTTACCATGGCGTACTACCATCACCATCACC-3`
Antisense primer:
5`-GCTACGAATTCTCAACTGATGGGGGCTCCAG-3`

pEAK8-mBsepEGFP: identical to pEAK8 mBsep vector but with an C-terminal fusion of enhanced green fluorescent protein to locate the protein.
mBsep:
Sense primer:
5`-GCTACAAGCTTACCATGGCGTACTACCATCACCATCACC-3`
Antisense primer:
5`-GCTACGAATTCACTGATGGGGGCTCCAG-3`

EGFP:
Sense primer:
5`-GCTACGAATTCATGGTGAGCAAGGGCGAGG-3`
Antisense primer:
5`-GCATTAGCGGCCGCTTACTTGTAACAGCTCGTCCA-3`

Sequencing primer for the mBsep constructs:
bs1: 5`-GCTGATGTTTATGGGAAGTG-3`
bs2: 5`-AGGTTGTGGGTAATCACTGG-3`
bs3: 5`-TGAGTGTGGCCAAGTTCACG-3`
bs4: 5`-GATTTTCCTCTGTGTCATAA-3`
bs5: 5`-AGCACAGCACTACAGCTCAT-3`
bs6: 5`-GACAGAAGCAAAGGGTAGCC-3`
bs7: 5`-TACTCACAAAGAAACAGGCA-3`
bs8: 5`-TTTGTGTGCAGCCATTAATG-3`
bs9: 5`-CAGATGCTTCCCAAGTTCAA-3`
bs10: 5`-TGAGGTTGAGCTGGAGAAGT-3`
bs11: 5`-GTACAGTGGAGCAGGTGAAA-3`

Constructs of prokaryotic ABC Transporter

Thermotoga maritima

TM0043:
Sense primer:
5`-GGTGGATCCGTGATTTTGAGTCCGGAGAAGTTTCTGC-3`
Antisense primer:
5`-GGCTACTCGAGCGCATTTTCGAACTGTATCTTGTAAAGG-3`

TM0287:
Sense primer:
5`-GCTCAGAGCTCATGAAAACACTCGCCAGATATTTAAAGCC-3`
Antisense primer:
5`-GCTCACAGCGGCCGCTGGCATCGTTCATCACCCCG-3`
Antisense primer with no tag:
5`-GCTCACAGCGGCCGCTCAGGCATCGTTCATCACCCCGT-3`
Nucelotide Binding domain:
Sequencing primer (start at 600)
5`-GGTGTCAGGGTAGTGAGGGC-3`

TM0288:
Sense primer:
5`-GCTGGATCCATGCCTGAGATCAGAAGAAGGCC-3`
Antisense primer:
5`-GCTACCTCGAGTTCTTTTTCTACAACGAGACCGTATTGGC-3`
Antisense primer without tag:
5`-GCTACCTCGAGtcaTTCTTTTTCTACAACGAGAC-3`
antisense primer: only avi-tag:
5`-GCTGACTCGAGTCACTCGACTTCGTGCCATTCGATTTTCTG-3`
Sequencing primer (start at base 596)
5`-GCCAAACCAGAAAATACTTCTACG-3`

TM1256:
Sense primer:
5`-GCTGGATCCTTGAAGACTTCCTCAATTCTCGTTGATTTC-3`
Antisense primer:
5`-GCGACTGAAGCGGCCGCTTCCCTCCTCGATGAGCTGTTTTCTG-3`

TM1310:
Sense primer:
5`-GCTGGATCCATGAATCTCAGGAGTATCCAG-3`
Antisense primer:
5`-GCTCACTCGAGGACATCTTTATCAGCGATCC-3`

TM1319:
Sense primer:
5`-GCTGGATCCATGAAAACTGGAAAGTTTCGGCTCTTGAG-3`
Antisense primer:
5`-GCGACTGGCGGCCGCTTACCTCCAATTGACTCCTTACAATCTC-3`

TM1328:
Sense primer:
5`-GCTGGATCCATGAAATTTTTCTCAGCAAGTAACGTCAGAG-3`
Antisense primer:
5`-GCTCACTCGAGTCTCATCAGCTGGCTCTTCA-3`

Escherichia coli:

D-galactose transporter MglC/A:
Mgla:
Sense primer:
5`-GCTAGGGATCCATGGTCAGCTCAACGACTCC-3`
Antisense primer:
5`-GCTACGTGCGGCCGCAGGTGCAAAGACGCAAGACG-3`

MglC:
Sense primer:
5`-GCTAGGGATCCATGAGTGCGTTAAATAAGAA-3`
Antisense primer:
5`-GCTACCTCGAGTTTCTTACGCGCGTATTTCA-3`

CVAB
1:
Sense primer:
5`-GATCCATGACTAACAGGAATTTCAGACAAATTATA-3`
Antisense primer:
5`-TCGAGAATAGAAATAACTCTATCAACAGTTCTCAA-3`

2:
Sense primer:
5`-CATGACTAACAGGAATTTCAGACAAATTATA-3`
Antisense primer:
5`-GAATAGAAATAACTCTATCAACAGTTCTCAA-3`

HLYB
1:
Sense primer:
5`-GATCCATGGATTCTTGTCATAAAATTGATTATGGG-3`
Antisense primer:
5`-TCGAGGTCTGACTGTAACTGATATAAGTAACTGTA-3`

2:
Sense primer:
5`-CATGGATTCTTGTCATAAAATTGATTATGGG-3`
Antisense primer:
5`-GGTCTGACTGTAACTGATATAAGTAACTGTA-3`

Sense primer:
5`-ATGGATTCTTGTCATAAAATTGATTATGGG-3`
Antisense primer:
5`-GTCTGACTGTAACTGATATAAGTAACTGTA-3`

MACB
1:
Sense primer:
5`-GATCCATGACGCCTTTGCTCGAATTAAAGGATATT-3`
Antisense primer:
5`-TCGAGCTCTCGTGCCAGAGCATCTACTGGATCCAG-3`

2:
Sense primer:
5`-CATGACGCCTTTGCTCGAATTAAAGGATATT-3`
Antisense primer:
5`-GCTCTCGTGCCAGAGCATCTACTGGATCCAG-3`

YBIZ
1:
Sense primer:
5`-GATCCATGACGCCTTTGCTCGAATTAAAGGATATT-3`
Antisense primer:
5`-TCGAGCTCTCGTGCCAGAGCATCTACTGGATCCAG-3`

2:
Sense primer:
5`-CATGACGCCTTTGCTCGAATTAAAGGATATT-3`
Antisense primer:
5`-GCTCTCGTGCCAGAGCATCTACTGGATCCAG-3`

YDDA:

1: Sense primer:
5`-GATCCATGCATAACGACAAAGATCTCTCTACGTGG-3`
Antisense primer:
5`-TCGAGTTGGCCAAACTGCATTTTGTGAAGTTGC-3`

2:
Sense primer:
5`-C ATGCATAACGACAAAGATCTCTCTACGTGG-3`
Antisense primer:
5`-GTTGGCCAAACTGCATTTTGTGAAGTTGC-3`
ydda_pET_28_ch_neu_primer
Sense primer:
5`-GCTCAGGATCCATGAAATCGCGCCAG-3`
Antisense primer:
5`-GCTATCTCGAGTAAAACCGCGCTAAT-3`

Yoji:

1:
Sense primer:
5`-GATCCATGGAACTTCTTGTACTTGTCTGGCGGCAG-3`
Antisense primer:
5`-TCGAGTGCCGTCCGGGCAACGGCATCACGCGAAGC-3`

2:
Sense primer:
5`-CATGGAACTTCTTGTACTTGTCTGGCGGCAG-3`
Antisense primer:
5`-GTGCCGTCCGGGCAACGGCATCACGCGAAGC-3`

yoji pET28_ch_neu_primer:
Sense primer:
5`-ATGGAACTTCTTGTACTTGTCTGGCGGCAG-3`
Antisense primer:
5`-TGCCGTCCGGGCAACGGCATCACGCGAAGC-3`

Zinc transporter ZnuB/C:

znuB:

Sense primer:
5`-GCTAGGGATCCATGATTGAATTATTATTTCC-3`
Antisense primer:
5`-GCTAGCTCGAGGCTGGCCTGCTTTTTCATCA-3`

znuC:

Sense primer:
5`-GCTAGGGATCCATGACAAGTCTGGTTTCCCT-3`
Antisense primer:
5`-GCTAGCTCGAGTGAGCGATCATTTCCCCGAC-3`

Arabinose transporter: AraH/G:
araG:
Sense primer:
5`-GCTAGGGATCCATGCAACAGTCTACCCCGTA-3`
Antisense primer:
5`-GCTAGCTCGAGGGCAACAGCCTGGCTGACTT-3`

araH:
Sense primer:
5`-GCTAGGGATCCATGATGTCTTCTGTTTCTAC-3`
Antisense primer:
5`-GCTAGCTCGAGGACAGTGCGTTTCGCTTTTT-3`

D-xylose transporter XylH/G
xylG:
Sense primer:
5`-GCTAGGGATCCATGCCTTATCTACTTGAAAT-3`
Antisense primer:
5`-GCTAGCTCGAGGACGGATTGCTTTTCGACAT-3`
xylH:
Sense primer:
5`-GCTAGGGATCCATGTCGAAAAGCAATCCGTC-3`
Antisense primer:
5`-GCTAGCTCGAGAGAACGGCGTTTGGTTGCGG-3`

Glutamine transporter GlnP/Q:
glnP:
Sense primer:
5`-GCTAGGGATCCATGCAGTTTGACTGGAGTGC-3`
Antisense primer:
5`-GCTAGCTCGAGCAGGATTTTCATCCTTCTTT-3`
glnQ:
Sense primer:
5`-GCTACGGATCCATGATTGAATTTAAAAACGTC-3`
Antisense primer:
5`-GCTACCTCGAGAGAGACGTGCTGCAAAAATT-3`

Taurine transporter: TauC/B
tauB:
Sense primer:
5`-GCTAGGGATCCATGCTGCAAATCTCTCATCT-3`
Antisense primer:
5`-GCTAGCTCGAGTGAGAACGCCTCCCGTTGCT-3`
tauC:
Sense primer:
5`-GCATCGAATTCATGAGTGTGCTCATTAATGA-3`
Antisense primer:
5`-GCTAGCTCGAGTTGTACTTCTCCATGCCAGG-3`

Macroliden transporter macB:
Sense primer:
5`-ATGACGCCTTTGCTCGAATTAAAGGATATT-3`
Antisense primer:
5`-CTCTCGTGCCAGAGCATCTACTGGATCCAG-3`

Lipid flipase MsbA:
Sense primer:
5`-ATGCATAACGACAAAGATCTCTCTACGTGG-3`
Antisense primer:
5`-TTGGCCAAACTGCATTTTGTGAAGTTGCGC-3`

Pyrococcus horikoshii

PH1735:
Sense primer:
5`-GCTGAGCTCGTGAGGAATATGAGAAATAACG-3`
Antisense primer:
5`- GCGATCTCGAGAACTTCCTGAACTTGAGAGG-3`

PH1736:
Sense primer:
5`-GCTGAATTCATGGATGGTTCGCTAAGGC-3`
Antisense primer:
5`-GCTCACTCGAGTTTCTCATATTCCTCACCTATAGC-3`

Helicobacter pylori

HP 0600:
Sense primer:
5`-GCTGGATCCATGCAAACACCAATGGATAC-3`
Antisense primer:
5`-GCTACCTCGAGGTATTGCTGCTTCAAAAACA-3`

HP1082:
Sense primer:
5`-GCTGAATTCTTGAAACTCTTTTTCAAGCG-3`
Antisense primer:
5`-GCTACCTCGAGGCATTCTGTCAAACGC-3`

Bacillus subtilis:

YFIB:
Sense primer:
5`-GCTGGATCCATGCAAAAGGCTTTGTCGTTTTTAAAACC-3`
Antisense primer:
5`-GCTCACTCGAGGCATGACTCGCTCCCCTCCCTTC-3`

YFIC:
Sense primer:
5`-GCTGGATCCATGCTAAAAGACATCCGCAAGCCTTTCC-3`
Antisense primer:
5`-GCTCACTCGAGTTTCTCAAATTGGCTTTCATATAAGTCACTG-3`

YHEH:
Sense primer:
5`-GCTGGATCCATGAAAATAGGAAAAACGTTATGGAGATACGC-3`
Antisense primer:
5`-GCTCACTCGAGTGCAATGGAATGTTTCTGTCCCTTTTGC-3`

YHEI:
Sense primer:
5`-GCTGGATCCATGTTTTCAGTTTTGAAAAAGCTTGGCTGG-3`
Antisense primer:
5`-GCTCACTCGAGTGCCCCTGCTCCCCCTTCTTCC-3`

YKNU:
Sense primer:
5`-GCTGGATCCATGGAGACATTTAAAAGATTAAAAATGTATTATTGGC-3`
Antisense primer:
5`-GCTCACTCGAGTCCGACCTCATGCGGCTC-3`

YKNV:
Sense primer:
5`-GCTGGATCCATGAAACAAGCGAAAAAACAGGGGGTTTTAG -3`
Antisense primer:
5`-GCTCACTCGAGTTCTATCGCTGTGCTGTATTGCGCTTTC-3`

YWJA:
Sense primer:
5`-GCTGGATCCGTGCTGCGGCAATTTTTTTCATATTATAAACC -3`
Antisense primer:
5`-GCTCACTCGAGTCGATGCACCATTTGCCCGAATTG-3`

Table 8: Expression Overview

Constructs	BL21(DE3)	BL21(DE3) cRILplus	BL21(DE3) pLysS	C41	C43
1st round of selection					
pET 20 constructs					
TM 0043 pET20 NH	-	-	-+		
TM 0043 pET20 CH	+-	-	-+		
TM 0287 pET20 NH	-	-	+		
TM 0287 pET20 CH	-	+	+		
TM 0288 pET20 NH	-	-	+		
TM 0288 pET20 CH	+	++	+		
TM 1256 pET20 NH	-	-	-		
TM 1256 pET20 CH	-	-	-		
TM 1310 pET20 CH	-	-	-		
TM 1319 pET20 NH	-	-	-		
TM 1319 pET20 CH	-	-	-		
TM 1328 pET20 NH					
TM 1328 pET20 CH	-	-	+-		
HP 0600 pET20 NH	-	-	-		
HP 0600 pET20 CH	-	-	-		
PH 1735 pET20 NH	-	-	-		
PH 1735 pET20 CH	-	-	-		

Table 8: continued

PH 1736 pET20 NH	-	-	-		
PH 1736 pET20 CH	-	-	-		
YWJA pET20 NH					
YWJA pET20 CH					
Constructs with pelB leader sequence (frame)					
TM 1256 pET20 frame	-	-	-		
TM 1319 pET20 frame	-	-	-		
TM 1328 pET20 frame	-	-	-		
PH 1735 pET20 frame					
PH 1736 pET20 frame					
HP 1082 pET20 frame					
HP 1082 pET20 CH NH	+	+	+		
pET 21/28 constructs					
TM 0043 pET21 CH					
TM 0043 pET21 frame					
TM 0043 pET28 CH	+	+	-		
TM 0043 pET28 frame					
TM 287 pET21 CH	+++	++		+	+
TM 287 pET21 frame	-				
TM 287 pET28 CH	0	0	0		
TM 287 pET28 frame	0	0	0		
TM 288 pET21 CH	0	0	0		
TM 288 pET21 frame	0	0	0		
TM 288 pET28 CH	+++	+++	-	+	+
TM 288 pET28 frame	0	0	0		
TM 1256 pET21 CH	0	0	0		
TM 1256 pET21 frame	0	0	0		
TM 1256 pET28 CH	+++	+++	+		
TM 1256 pET28 frame	-	-	-		
TM 1319 pET21 CH					
TM 1319 pET21 frame					
TM 1319 pET28 CH	-	-	-		
TM 1319 pET28 frame	-	-	-		
TM 1328 pET21 CH					
TM 1328 pET21 frame					
TM 1328 pET28 CH	+++	-	-		
TM 1328 pET28 frame	-	-	-		
PH 1735 pET21 CH					
PH 1735 pET21 frame					
PH 1735 pET28 CH	+++	+++	+		
PH 1735 pET28 frame	++	-	-		

Table 8: continued

PH 1736 pET21 CH					
PH 1736 pET21 frame					
PH 1736 pET28 CH	+	+	-		
PH 1736 pET28 frame	-	-	-		
HP 1082 pET21 CH					
HP 1082 pET21 frame					
HP 1082 pET28 CH	++	+	-		
HP 1082 pET28 frame	-	-	-		
2nd round of selection					
Zink					
B pET21 CH	+			-	-
B pET21 NH					
B pET28 CH					
B pET28 NH					
C pET21 CH -ATP-	++			-	-
C pET21 NH	+-				
C pET28 CH	++				
C pET28 NH	+-				
Taurine					
B pET21 CH -ATP-					
B pET21 NH					
B pET28 CH					
B pET28 NH	++			+	-
C pET21 CH	++			++	-
C pET21 NH					
C pET28 CH					
C pET28 NH					
Arabinose					
G pET21 CH -ATP-	++			+	-
G pET21 NH					
G pET28 CH					
G pET28 NH					
H pET21 CH					
H pET21 NH					
H pET28 CH	+			+-	-
H pET28 NH	+-				
Xylose	+-				
G pET21 CH -ATP-	-			-	-
G pET21 NH	-			-	-
G pET28 CH	++			-	-
G pET28 NH	0				
H pET21 CH	-			-	-
H pET21 NH	+			-	-
H pET28 CH	-			-	-
H pET28 NH	0				
Glutamine					
P pET21 CH					
P pET21 NH	+			-	-

Table 8: continued

P pET28 CH	+		
P pET28 NH			
Q pET21 CH -ATP-			
Q pET21 NH	+		
Q pET28 CH			
Q pET28 NH	+	-	-
3rd round of selection			
HLYB			
MACB			
MSBA			
YBIZ			
YDDA	++		
YOJI	++		

-: no expression; 0: very weak expression, +: weak expression; ++: medium expression; +++: strong expression.

F.4 CitS

Expression and Purification of CitS

The CitS was expressed using freshly transformed C43 *E.coli* strains. A single colony was used to inoculate a 10 ml LB culture, supplemented with 1% glucose and ampicillin, and was grown at 30° C overnight. The media of the overnight preculture was removed by centrifugation; the cells were resuspended in 4 ml buffered TB media (24 g/l yeast extract, 12 g/l tryptone peptone, 3.6 ml/l glycerol, 12.6 g/l K_2HPO_4 and 2.3 g/l KH_2PO_4) and used to inoculate 500 ml TB media. The cells were grown to an OD_{600} of 0.6-0.7 at 30°C, cooled down to 18°C in an ice bath for 15-20 min and protein expression was started by adding 1.4 mM IPTG and the cells were incubated at 25°C for 18 hours. The cells were harvested by centrifugation and suspended with twice the amount of lysis buffer (20 mM Tris, 300 mM NaCl, 20 mM NaH_2PO_4, pH 7.5) and incubated one hour with lysozyme (1mg /ml), DNAse and RNAse. Cells were disrupted by sonification and three passes of FrenchPress.

To solubilize the protein, 2 % (w/v) dodecyl β-D-maltopyranoside (DDM) was added to the crude extract and the solution was gently shaken for 2 hours at 4°. The suspension was centrifuged 30 min at 40000 x g to remove non solubilized material. The solution was filtered through a 0.22 µm filter and, prior to loading it to an equilibrated Ni-NTA affinity column (10 mM Tris, 300 mM NaCl, 0.1% DDM, pH 7.4), the pH was adjusted to 8.0 with 0.5 M NaOH. The column was washed with 10 column volumes of buffer 1 (10 mM Tris, 300 mM NaCl, 0.1% DDM, pH 7.4), 10 column volumes of buffer 2 (10 mM Tris, 1M NaCl, 0.1% DDM, pH 7.4), 10 column volumes of buffer 3 (10 mM Tris, 500 mm NaCl, 100 mM Imidazole, 0.1% DDM, pH 7.4) and finally the protein was eluted with 3-5 column volumes of elution buffer (10 mM Tris, 300 mm NaCl, 500 mM Imidazole 0.1% DDM, pH 7.4).

The eluted protein was concentrated to 500 µl and applied to an equilibrated size exclusion chromatography column (Superose 6 23 ml; 10 mM Tris, 150 mM NaCl, 0.05% DDM, pH 7.4). Fractions containing CitS were pooled, concentrated to 10-30 mg /ml and used for crystallization experiments.

Initial Crystallization of CitS

Initial screening of crystallization conditions was carried out on a small scale using the NCCR crystallization facility at the University of Zürich. 100 µl protein solution was mixed with 100 µl reservoir solution in 96 well plates and incubated at 4°. The crystallization trials were inspected immediately after set up and after 1, 3, 7 and 10 days of incubation. After 10 days the wells were inspected every 2 weeks for a total of 4 months from the initial set up of the plate. The initial screens included pH ranges of 4.5-9.4 in steps of 1 pH unit, poly ethylene glycol with a mean molecular weight of 400 or 4000 and the following salt conditions:

150 mM or 1 M NaCl
200 mM $(NH_4)_2SO_4$ or 50 mM Mg-acetate
100 or 200 mM K_3PO_4 or 50 or 100 mM of each of Li_2SO_4 and Na_2SO_4
100 mM KCl or 100 mM $CaCl_2$
1 M ammonium formate or 50 mM Zn-acetate or 100 mM KI
100 mM K-Na-tartrate

Table 9: Heavy Metal Derivatives

Heavy metal Crystal number	Amount (mM) Time[1]	Co-crystallized (mM)	Fluorescence signal	Resolution (Å)	Anomalous signal[2]	R-factor[3] (%)
Native	-	-	-	3.6	-	-
CsCl 37	-	100	nd	6.0	1.20	28.4
$KOsCl_4$ 23-1	1 30m/30m	-	nd	7.0	1.08	11.2
$K_2Pt(CN)_4$ 1041	1 1m/-	-	✓	6.0	1.00	29.1
$Pb(NO_3)_2$ 1043	1 1m/-	-	✓	6.0	1.00	21.0
TaBr 1275	2 1h/1m	-	✓	6.0	1.12	22.7
$(H_2NCH_2CH_2NH_2)PtCl_2$ 1280	2 1h/1m	-	✓	5.6	1.16	14.4
$(H_2NCH_2CH_2NH_2)PtCl_2$ 1281	2 1h/1m	-	✓	5.25	1.10	13.0
Thiomersal 1321	0.05 72h/1m	-	Nd	7.3	1.18	32.1
$Hg(CH_3CO_2)_2$ 1337	-	0.01	Nd	6.9	1.01	26.5
Ta_6Br_{14} 1424	0.5 24h/2m	-	✓	7.0	1.00	nd
Thiomersal 1438	-	0.01	Nd	7.0	1.11	34.2
$HoCl_3$ 1444	-	0.01	Nd	5.5	1.10	16.4
$HoCl_3$ 1446	-	0.01	Nd	6.0	1.14	19.9
K_2PtCl_6 1678	-	0.1	Nd	5.0	1.05	12.0
$KAu(CN)_2$ 1679	-	0.1	Nd	5.0	1.30	17.5
Thiomersal 1680	-	0.1	nd	5.0	1.0	15.0
$Pb(NO_3)_2$ 1704	1 5m/-	-	✓	6.9	1.1	Nd
Dichloro (2,2',6',2" ter-pyridine) platinum (II) 1710	1 5m/1m	-	✓	6.9	1.0	Nd
$Hg(CH_3)Cl$ 1714	2 5m/-	-	✓	7.0	1.15	Nd
$Hg(CH_3)Cl$ 1717	2 5m/-	-	✓	7.0	1.1	Nd
$Hg(CH_3)Cl$ 1802	0.5 1h/20m	-	✓	7.0	1.0	Nd
Hg(CH3)Cl 1804	1 1h/20m	-	Nd	7.3	1.0	Nd
$HgCl_2$ 1813	2 20m/20m	-	✓	6.5	1.0	Nd
$HgCl_2$ 1818	1 40m/20m	-	✓	5.5	1.05	Nd
K_2PtCl_6 1830	1 40m/20m	-	✓	6.0	1.00	Nd
K_2PtCl_6 1834	2 40m/20m	-	✓	6.8	1.00	nd

[1] time of soaking and back soaking (m minutes, h hours, d days); [2] Ratio of $\square_{norm}/\square_{anomalous}$ determined by XSCALE (Kabsch, 1993); [3] compared to the native data set;

F.5 β-Catenin

β-Catenin is involved in cell adhesion, transcription and cell signaling in the wingless (WNT) pathway; therefore the strict control of its availability during normal development is crucial. Inappropriate activation of β-catenin is related to several cancer types. The aim of the project was to purify and co-crystallize a fragment of β-catenin with several peptides shown to be involved in a nuclear complex between β-catenin and TCF.

Introduction

The protein β-catenin consists of 12 armadillo repeats composed of 42 amino acids that are flanked by a 150 amino acid N-terminal domain and a 100 amino acid C-terminal domain. It is found in several complexes including the nuclear complex with the transcription factors T-cell factor (TCF), legless and pygopus (van Es *et al.*, 2003, Kramps *et al.*, 2002), the complex with cadherin and α-catenin involved in cell adhesion (Perez-Moreno *et al.*, 2003) and a complex involving the proteins Axin, the adenomatous polyposis coli protein (APC) and the two kinases glycogen synthase kinase-3 (GSK3) and casein kinase I (CK I) (Rubinfeld *et al.*, 1993, Behrens *et al.*, 1998). The molecules axin, APC, TCF and cadherins bind to overlapping but not identical sites within the Armadillo repeats of β-catenin (Figure 74) whereas the complex of α- and β- catenin uses interactions formed in the N-terminal domain of the protein (Huber *et al.*, 1997).

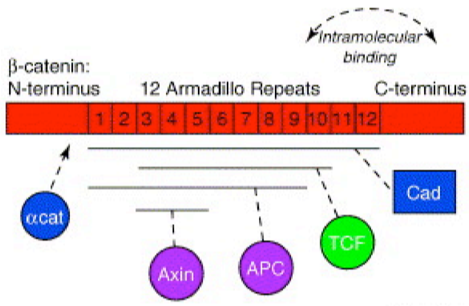

Figure 74: Interaction sites of β-catenin. α-Catenin interacts with the N-terminal domain of β-catenin whereas the proteins Axin, APC, TCF, E-cadherin and the C-terminus of β-catenin bind to the central 12 armadillo repeats (Harris & Peifer, 2005).

Results of β-catenin

Expression

The construct used for expression consisted of amino acids 134-664 of the naturally occurring β-catenin protein. This fragment includes the 12 armadillo repeats and 17 amino acids of the N terminal domain. It was expressed as C-terminal fusion protein to GST with a thrombin cleavage site. The construct was obtained by Sandipan Chatterjee of the group of Prof. Konrad Basler of the University of Zürich.
A freshly transformed BL21 (DE3) colony was inoculated o/n in 5 ml LB media containing Ampicillin (100µg/ml) and grown at 37 °C. 500 ml LB media were inoculated with the o/n culture grown to an OD_{600} of 0.5 and induced with 0.2 mM IPTG. After 3h the cells were harvested by centrifugation (30 min, 3500*g, 4°C) and frozen at -20°.

Purification

A 3 g frozen cell pellet was resuspended in 20 ml lysis buffer (50 mM NaCl, 50 mM Tris, 2mM DTT, pH 7.4) and DNAse and protease inhibitors were added. The cell suspension was lyzed by passing the suspension twice through a FrenchPress (1200 psi, 4°C) and centrifuged (30 min, 40000*g, 4°C) to pellet the remaining cell debris. The supernatant was applied to 5 ml GST Sepharose FF (Pharmacia) equilibrated in lysis buffer. The suspension was gently shaken at 4°C for 90 minutes. The suspension was applied to an empty column, washed with 100 ml lysis buffer and the protein was eluted with 20 ml elution buffer (25 mM GSH, 50 mM Tris, 50 mM NaCl, pH 7.4). The purification process is shown in Figure 75, indicating the high purity of the GST-β-catenin fusion protein after the first purification step.

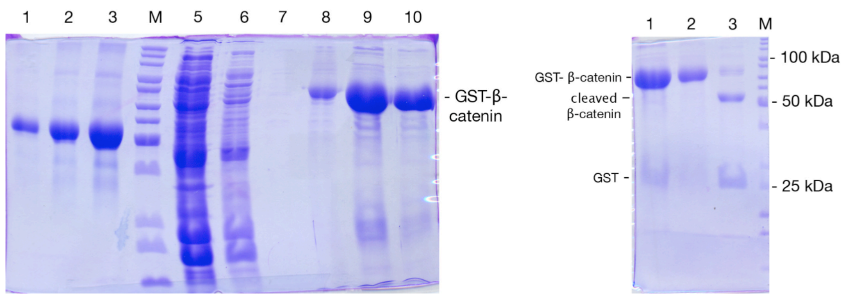

Figure 75: Purification and cleavage of GST-β-catenin. Left SDS-PAGE: lane 1: 2.5 µg BSA; lane 2: 5 µg BSA; lane 3: 10 µg BSA; lane 4: Marker proteins; lane 5: unbound sample of the GST-Sepharose column; lanes 6-7: start and end of the washing step; lanes 8-10: fractions of the elution with 25 mM GSH. Right SDS-PAGE: lane 1: elution without thrombin; lane 2: elution without thrombin (1/3 load); lane 3: digested protein after 18 h incubation at 20°C; lane 4: Marker proteins.

The fusion protein was cleaved by diluting the eluate to 50 ml with buffer A (25 mM Ethanolamine, 50 mM NaCl, 0.1% β-mercaptoethanol, pH 9.5) and incubation with 50 U Thrombin (Sigma) at 20°C o/n. Figure 75 illustrates the purity of the protein after cleavage. The three species uncleaved fusion protein, free β-catenin and free GST are present in the sample and had to be separated.

Anionic exchange chromatography was used to separate the GST and β-catenin (Figure 76). The digested solution was dialyzed against buffer A, applied to an 8 ml MonoQ column (Pharmacia) previously equilibrated in buffer A. The protein was eluted by a liner salt gradient of 0-170 mM NaCl. The shallow gradient enabled the separation of β-catenin, the uncleaved protein and the cleaved GST fusion protein. β-Catenin eluted at a conductivity of ~10 mS, the GST fusion protein at 17-18 mS whereas the uncleaved GST β-catenin protein was found to elute at around 15 mS. After SDS-PAGE analysis the fractions containing pure β-catenin were pooled and concentrated yielding approximately 1-2 mg pure β-catenin per liter of culture.

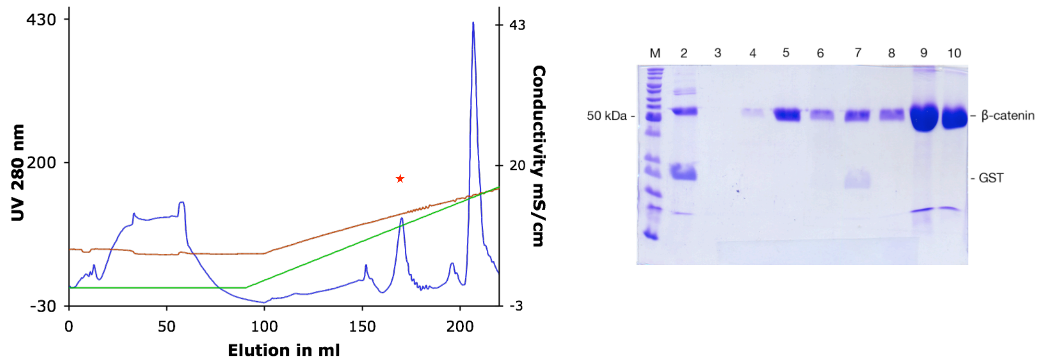

Figure 76: Anionic exchange chromatography and the corresponding SDS-PAGE. The absorption at 280 nm, the conductivity and the theoretical gradient are shown in blue, brown and green, respectively. A gradient of 0-170 mM NaCl in 17 column volumes was applied. Pure β-catenin elutes at ~ 10 mS (star). **SDS-PAGE**: lane 1: Marker proteins; lane 2: sample applied to the column; lane 3: unbound material; lane 4: peak at 150 ml; lane 5: peak at 170 ml; lane 6: peak at 195 ml; lane 7: peak at 215 ml; lane 8: merged fractions; lane 9: concentrated sample; lane 10: concentrated sample after filtration.

Peptides

Several β-catenin protein complexes are known and structurally characterized (Daniels *et al.*, 2001, Graham *et al.*, 2002, Graham *et al.*, 2001, Graham *et al.*, 2000, Huber & Weis, 2001). To gain insights into the binding of β-catenin to human legless protein, several peptides were synthesized and co-crystallized with purified β-catenin. Binding studies and GST pull down experiments of β-catenin and armadillo with these peptides revealed the minimal amino acids sequence required for a sufficient interaction and binding (Kramps *et al.*, 2002).

LP-1: legless basic peptide 1 (HD-2, amino acids 515-544)

N-term: ENLTPQQRQHREEQLAKIKKMNQFLFPENE

LP-2: BCL9 basic peptide 1 (HD-2) aa 349-378

N-term: DGLSQEQLEHRERSLQTLRDIQRMLFPDEK

LP-3: legless extended basic peptide 2 (HD-2) aa 510-549

N-term: VKVPDENLTPQQRQHREEQLAKIKKMNQFLFPENENSVGA

LP-4: BCL9 extended basic peptide 2 (HD-2) aa 344-383

N-term: LGENPDGLSQEQLEHRERSLQTLRDIQRMLFPDEKEFTGA

LP-6: human TCF-4 aa 1-53

N-term: MPQLNGGGGDDLGANDELISFKDEGEQEEKSSENSSAERDLADVKS SLVNESE

Peptides 1 and 3 are peptides derived of drosophila whereas peptides 2 and 4 are the human homologues of the product of the B cell lymphoma 9 gene. 5 amino acids were extended on both sides of the minimal sequence that showed binding to the target protein in the two peptides 2 and 4.

Crystallization of β-Catenin and β-Catenin peptide complexes

The protein was concentrated to 5 mg/ml, incubated with the different peptides in 1:1 or 1:3 ratio and crystallized in hanging or sitting drop plates. 1 µl of protein was mixed with 1µl of reservoir solution and stored over 900 µl reservoir solution at 20°C. Small crystals could be observed using 16-20% PEG 4000, 100 mM Tris pH 8.0, 200 mM NaCl and 10 mM DTT as precipitant and two peptides LP4 and LP6 in a 1+1.1 molar ratio. Although the crystals could be reproduced, it was not possible to increase their size, neither by screening for additives or seeding techniques. Different protein: peptide ratios and different combinations of LP6 and LP1-4 as binding partners of β-Catenin were screened, but none of the variations produced crystals that were suitable to structural studies.

Discussion

This part of the thesis establishes the expression and purification protocol of the protein β-catenin that has not previously been reported in literature. The low solubility of β-catenin around its pKa value of 8.4 required a GST containing fusion protein that prevented the aggregation of β-catenin. The GST-β-catenin fusion protein behaved well over a wide range of pH values and the first purification step yielded protein off sufficient amount and quality for biochemical studies (Kramps *et al.*, 2002). For structural studies the protein had to be further purified including the cleavage with thrombin at pH 9.5 and a subsequently separation of the cleavage products. The separation was carried out using anionic exchange chromatography at pH 9.5 yielding pure β-catenin. Crystallization trials with the various peptides revealed small crystals of 10 x 10 x10 μm^3 for the peptides LP-4 and LP-6. Notably, the addition of LP-6 to β-catenin, before or after cleavage with thrombin, increased the solubility and stability of the protein and no aggregation or precipitation, even at pH values around the pKa value of β-catenin, were observed. The behavior of β-catenin in complex with LP-6 (TCF4 peptide) might explain the various reported structures of β-catenin that were crystallized with this peptide (Daniels *et al.*, 2001, Graham *et al.*, 2001).

Chapter G - References

Abele, R. & Tampe, R. (2004). Physiology 19, 216-224.

Abramson, J., Smirnova, I., Kasho, V., Verner, G., Kaback, H. R. & Iwata, S. (2003). Science 301, 610-615.

Adams, P. D., Gopal, K., Grosse-Kunstleve, R. W., Hung, L. W., Ioerger, T. R., McCoy, A. J., Moriarty, N. W., Pai, R. K., Read, R. J., Romo, T. D., Sacchettini, J. C., Sauter, N. K., Storoni, L. C. & Terwilliger, T. C. (2004). Journal of Synchrotron Radiation 11, 53-55.

Adams, P. D., Pannu, N. S., Read, R. J. & Brunger, A. T. (1997). Proceedings of the National Academy of Sciences of the United States of America 94, 5018-5023.

Al-Sa'doni, H. & Ferro, A. (2000). Clinical Science 98, 507-520.

Ambudkar, S. V., Kim, I. W. & Sauna, Z. E. (2006). Eur J Pharm Sci 27, 392-400.

Arita, K., Hashimoto, H., Shimizu, T., Nakashima, K., Yamada, M. & Sato, M. (2004). Nature Structural & Molecular Biology 11, 777-783.

Bairoch, A., Apweiler, R., Wu, C. H., Barker, W. C., Boeckmann, B., Ferro, S., Gasteiger, E., Huang, H., Lopez, R., Magrane, M., Martin, M. J., Natale, D. A., O'Donovan, C., Redaschi, N. & Yeh, L. S. (2005). Nucleic Acids Res 33, D154-159.

Banatao, D. R., Cascio, D., Crowley, C. S., Fleissner, M. R., Tienson, H. L. & Yeates, T. O. (2006). Proc Natl Acad Sci U S A 103, 16230-16235.

Becker, A., Fritz-Wolf, K., Kabsch, W., Knappe, J., Schultz, S. & Volker Wagner, A. F. (1999). Nature Structural Biology 6, 969-975.

Bedford, M. T. & Richard, S. (2005). Molecular Cell 18, 263-272.

Behrens, J., Jerchow, B. A., Wurtele, M., Grimm, J., Asbrand, C., Wirtz, R., Kuhl, M., Wedlich, D. & Birchmeier, W. (1998). Science 280, 596-599.

Belrhali, H., Nollert, P., Royant, A., Menzel, C., Rosenbusch, J. P., Landau, E. M. & Pebay-Peyroula, E. (1999). Structure 7, 909-917.

Benabdelhak, H., Schmitt, L., Horn, C., Jumel, K., Blight, M. A. & Holland, I. B. (2005). Biochemical Journal 386, 489-495.

Binz, H. K., Amstutz, P., Kohl, A., Stumpp, M. T., Briand, C., Forrer, P., Grutter, M. G. & Pluckthun, A. (2004). Nature Biotechnology 22, 575-582.

Bogumil, R., Knipp, M., Fundel, S. M. & Vašák, M. (1998). Biochemistry 37, 4791-4798.

Bott, M., Meyer, M. & Dimroth, P. (1995). Molecular Microbiology 18, 533-546.

Boudker, O., Ryan, R. M., Yernool, D., Shimamoto, K. & Gouaux, E. (2007). Nature 445, 387-393.

Bruckdorfer, R. (2005). Mol Aspects Med 26, 3-31.

Brünger, A. T., Adams, P. D., Clore, G. M., DeLano, W. L., Gros, P., Grosse-Kunstleve, R. W., Jiang, J. S., Kuszewski, J., Nilges, M., Pannu, N. S., Read, R. J., Rice, L. M., Simonson, T. & Warren, G. L. (1998). Acta Crystallographica Section D-Biological Crystallography 54, 905-921.

Buchanan, S. K. (1999). Current opinion in structural biology 9, 455-461.

Burmeister, W. P. (2000). Acta crystallographica 56, 328-341.

Buss, K. A., Cooper, D. R., Ingram-Smith, C., Ferry, J. G., Sanders, D. A. & Hasson, M. S. (2001). Journal of Bacteriology 183, 680-686.

Callaghan, R., Berridge, G., Ferry, D. R. & Higgins, C. F. (1997). Biochimica et Biophysica Acta 1328, 109-124.

Chan, K. W., Zhang, H. & Logothetis, D. E. (2003). EMBO Journal 22, 3833-3843.

Chang, G. (2003). Journal of Molecular Biology 330, 419-430.

Chang, G. & Roth, C. B. (2001). Science 293, 1793-1800.

Chang, G., Roth, C. B., Reyes, C. L., Pornillos, O., Chen, Y. J. & Chen, A. P. (2006). Science 314, 1875.

Chang, G., Spencer, R. H., Lee, A. T., Barclay, M. T. & Rees, D. C. (1998). Science 282, 2220-2226.

Chen, J., Lu, G., Lin, J., Davidson, A. L. & Quiocho, F. A. (2003). Molecular Cell 12, 651-661.

Chimento, D. P., Mohanty, A. K., Kadner, R. J. & Wiener, M. C. (2003). Nature Structural Biology 10, 394-401.

Clamp, M., Cuff, J., Searle, S. M. & Barton, G. J. (2004). Bioinformatics 20, 426-427.

Collaborative Computational Project Number 4 (1994). Acta Crystallogr D Biol Crystallogr. 50, 760-763.

Daley, D. O., Rapp, M., Granseth, E., Melen, K., Drew, D. & von Heijne, G. (2005). Science 308, 1321-1323.

Daniels, D. L., Eklof Spink, K. & Weis, W. I. (2001). Trends Biochem Sci 26, 672-678.

Das, K., Butler, G. H., Kwiatkowski, V., Clark, A. D., Jr., Yadav, P. & Arnold, E. (2004). Structure (Camb) 12, 657-667.

Davidson, A. L., Laghaeian, S. S. & Mannering, D. E. (1996). Journal of Biological Chemistry 271, 4858-4863.

Davidson, A. L. & Sharma, S. (1997). J Bacteriol 179, 5458-5464.

Dawson, R. J. & Locher, K. P. (2006). Nature 443, 180-185.

Dean, M., Rzhetsky, A. & Allikmets, R. (2001). Genome Res 11, 1156-1166.

Diederichs, K., Diez, J., Greller, G., Muller, C., Breed, J., Schnell, C., Vonrhein, C., Boos, W. & Welte, W. (2000). EMBO Journal 19, 5951-5961.

Dimroth, P. & Thomer, A. (1990). Journal of Biological Chemistry 265, 7721-7724.

Drapier, J. C. & Hibbs, J. B., Jr. (1996). Methods in Enzymology 269, 26-36.

Dutzler, R., Campbell, E. B. & MacKinnon, R. (2003). Science 300, 108-112.

Enkvetchakul, D., Bhattacharyya, J., Jeliazkova, I., Groesbeck, D. K., Cukras, C. A. & Nichols, C. G. (2004). Journal of Biological Chemistry 279, 47076-47080.

Faller, M., Niederweis, M. & Schulz, G. E. (2004). Science 303, 1189-1192.

Ferguson, A. D., Chakraborty, R., Smith, B. S., Esser, L., van der Helm, D. & Deisenhofer, J. (2002). Science 295, 1715-1719.

Ferreira, K. N., Iverson, T. M., Maghlaoui, K., Barber, J. & Iwata, S. (2004). Science 303, 1831-1838.

Fritsche, E., Bergner, A., Humm, A., Piepersberg, W. & Huber, R. (1998). Biochemistry 37, 17664-17672.

Fromme, P., Jordan, P. & Krauss, N. (2001). Biochimica et Biophysica Acta 1507, 5-31.

Fu, D., Libson, A., Miercke, L. J., Weitzman, C., Nollert, P., Krucinski, J. & Stroud, R. M. (2000). Science 290, 481-486.

Galkin, A., Kulakova, L., Sarikaya, E., Lim, K., Howard, A. & Herzberg, O. (2004). Journal of Biological Chemistry 279, 14001-14008.

Garman, E. (2003). Current opinion in structural biology 13, 545-551.

Goodwin, D. C., Gunther, M. R., Hsi, L. C., Crews, B. C., Eling, T. E., Mason, R. P. & Marnett, L. J. (1998). Journal of Biological Chemistry 273, 8903-8909.

Gottesman, M. M. & Ambudkar, S. V. (2001). Journal of Bioenergetics & Biomembranes 33, 453-458.

Gow, A. J., Farkouh, C. R., Munson, D. A., Posencheg, M. A. & Ischiropoulos, H. (2004). American Journal of Physiology - Lung Cellular & Molecular Physiology 287, L262-268.

Graham, T. A., Clements, W. K., Kimelman, D. & Xu, W. (2002). Mol Cell 10, 563-571.

Graham, T. A., Ferkey, D. M., Mao, F., Kimelman, D. & Xu, W. (2001). Nat Struct Biol 8, 1048-1052.

Graham, T. A., Weaver, C., Mao, F., Kimelman, D. & Xu, W. (2000). Cell 103, 885-896.

Greller, G., Riek, R. & Boos, W. (2001). Eur J Biochem 268, 4011-4018.

Guan, L., Smirnova, I. N., Verner, G., Nagamori, S. & Kaback, H. R. (2006). Proceedings of the National Academy of Sciences of the United States of America 103, 1723-1726.

Guittet, O., Ducastel, B., Salem, J. S., Henry, Y., Rubin, H., Lemaire, G. & Lepoivre, M. (1998). Journal of Biological Chemistry 273, 22136-22144.

Haendeler, J., Hoffmann, J., Tischler, V., Berk, B. C., Zeiher, A. M. & Dimmeler, S. (2002). Nature Cell Biology 4, 743-749.

Hanafy, K. A., Krumenacker, J. S. & Murad, F. (2001). Medical Science Monitor 7, 801-819.

Harris, T. J. & Peifer, M. (2005). Trends in Cell Biology 15, 234-237.

Heinz, D. W. & Matthews, B. W. (1994). Protein Eng 7, 301-307.

Hessa, T., Kim, H., Bihlmaier, K., Lundin, C., Boekel, J., Andersson, H., Nilsson, I., White, S. H. & von Heijne, G. (2005). Nature 433, 377-381.

Heymann, J. B. & Engel, A. (2000). Journal of Molecular Biology 295, 1039-1053.

Higgins, C. F., Callaghan, R., Linton, K. J., Rosenberg, M. F. & Ford, R. C. (1997). Seminars in Cancer Biology 8, 135-142.

Higgins, C. F., Hiles, I. D., Salmond, G. P., Gill, D. R., Downie, J. A., Evans, I. J., Holland, I. B., Gray, L., Buckel, S. D. & Bell, A. W. (1986). Nature 323, 448-450.

Higgins, C. F. & Linton, K. J. (2004). Nature Structural & Molecular Biology 11, 918-926.

Hollenstein, K., Frei, D. C. & Locher, K. P. (2007). Nature 446, 213-216.

Huang, Y., Lemieux, M. J., Song, J., Auer, M. & Wang, D. N. (2003). Science 301, 616-620.

Huber, A. H., Nelson, W. J. & Weis, W. I. (1997). Cell 90, 871-882.

Huber, A. H. & Weis, W. I. (2001). Cell 105, 391-402.

Huber, T., Steiner, D., Röthlisberger, D. & Plückthun, A. (2007). J Struct Biol.

Humm, A., Fritsche, E., Steinbacher, S. & Huber, R. (1997). EMBO Journal 16, 3373-3385.

Hung, L. W., Wang, I. X., Nikaido, K., Liu, P. Q., Ames, G. F. & Kim, S. H. (1998). Nature 396, 703-707.

Hunte, C., Koepke, J., Lange, C., Rossmanith, T. & Michel, H. (2000). Structure 8, 669-684.

Hunte, C. & Michel, H. (2002). Current opinion in structural biology 12, 503-508.

Huopio, H., Vauhkonen, I., Komulainen, J., Niskanen, L., Otonkoski, T. & Laakso, M. (2002). Diabetes Care 25, 101-106.

Igarashi, Y., Aoki, K. F., Mamitsuka, H., Kuma, K. & Kanehisa, M. (2004). Molecular Biology & Evolution 21, 2149-2160.

Iwata, S., Ostermeier, C., Ludwig, B. & Michel, H. (1995). Nature 376, 660-669.

Iyer, P. P., Lawrence, S. H., Luther, K. B., Rajashankar, K. R., Yennawar, H. P., Ferry, J. G. & Schindelin, H. (2004). Structure 12, 559-567.

Janas, E., Hofacker, M., Chen, M., Gompf, S., van der Does, C. & Tampe, R. (2003). Journal of Biological Chemistry 278, 26862-26869.

Jordan, P., Fromme, P., Witt, H. T., Klukas, O., Saenger, W. & Krauss, N. (2001). Nature 411, 909-917.

Kabsch, W. (1993). Journal of Applied Crystallography 26, 795-800.

Karttunen, J. T., Lehner, P. J., Gupta, S. S., Hewitt, E. W. & Cresswell, P. (2001). Proceedings of the National Academy of Sciences of the United States of America 98, 7431-7436.

Kästner, C. N., Dimroth, P. & Pos, K. M. (2000). Archives of Microbiology 174, 67-73.

Kästner, C. N., Prummer, M., Sick, B., Renn, A., Wild, U. P. & Dimroth, P. (2003). Biophysical Journal 84, 1651-1659.

Kästner, C. N., Schneider, K., Dimroth, P. & Pos, K. M. (2002). Archives of Microbiology 177, 500-506.

Kaufmann, M., Lindner, P., Honegger, A., Blank, K., Tschopp, M., Capitani, G., Pluckthun, A. & Grutter, M. G. (2002). Journal of Molecular Biology 318, 135-147.

Kimoto, M., Tsuji, H., Ogawa, T. & Sasaoka, K. (1993). Archives of Biochemistry & Biophysics 300, 657-662.

Knappik, A., Ge, L., Honegger, A., Pack, P., Fischer, M., Wellnhofer, G., Hoess, A., Wolle, J., Plückthun, A. & Virnekäs, B. (2000). Journal of Molecular Biology 296, 57-86.

Ko, Y. H. & Pedersen, P. L. (2001). Journal of Bioenergetics & Biomembranes 33, 513-521.

Kohl, A., Binz, H. K., Forrer, P., Stumpp, M. T., Plückthun, A. & Grütter, M. G. (2003). Proceedings of the National Academy of Sciences of the United States of America 100, 1700-1705.

Kovari, L. C., Momany, C. & Rossmann, M. G. (1995). Structure 3, 1291-1293.

Kramps, T., Peter, O., Brunner, E., Nellen, D., Froesch, B., Chatterjee, S., Murone, M., Zullig, S. & Basler, K. (2002). Cell 109, 47-60.

Kuo, A., Gulbis, J. M., Antcliff, J. F., Rahman, T., Lowe, E. D., Zimmer, J., Cuthbertson, J., Ashcroft, F. M., Ezaki, T. & Doyle, D. A. (2003). Science 300, 1922-1926.

Landau, E. M. & Rosenbusch, J. P. (1996). Proceedings of the National Academy of Sciences of the United States of America 93, 14532-14535.

Leiper, J., Murray-Rust, J., McDonald, N. & Vallance, P. (2002). Proceedings of the National Academy of Science USA 99, 13527-13532.

Leiper, J. M., Santa Maria, J., Chubb, A., MacAllister, R. J., Charles, I. G., Whitley, G. S. J. & Vallance, P. (1999). Biochemical Journal 343, 209-214.

Lewis, H. A., Buchanan, S. G., Burley, S. K., Conners, K., Dickey, M., Dorwart, M., Fowler, R., Gao, X., Guggino, W. B., Hendrickson, W. A., Hunt, J. F., Kearins, M. C., Lorimer, D., Maloney, P. C., Post, K. W., Rajashankar, K. R., Rutter, M. E., Sauder, J. M., Shriver, S., Thibodeau, P. H., Thomas, P. J., Zhang, M., Zhao, X. & Emtage, S. (2004). EMBO Journal 23, 282-293.

Linton, K. J. & Higgins, C. F. (1998). Molecular Microbiology 28, 5-13.

Linton, K. J. & Higgins, C. F. (2002). Mol Membr Biol 19, 51-58.

Locher, K. P. (2004). Current opinion in structural biology 14, 426-431.

Locher, K. P., Bass, R. B. & Rees, D. C. (2003). Science 301, 603-604.

Locher, K. P. & Borths, E. (2004). FEBS Letters 564, 264-268.

Locher, K. P., Lee, A. T. & Rees, D. C. (2002). Science 296, 1091-1098.

Locher, K. P., Rees, B., Koebnik, R., Mitschler, A., Moulinier, L., Rosenbusch, J. P. & Moras, D. (1998). Cell 95, 771-778.

Lolkema, J. S., Enequist, H. & van der Rest, M. E. (1994). European Journal of Biochemistry 220, 469-475.

Lolkema, J. S., Sobczak, I. & Slotboom, D. J. (2005). FEBS Journal 272, 2334-2344.

Loll, B., Kern, J., Saenger, W., Zouni, A. & Biesiadka, J. (2005). Nature 438, 1040-1044.

Luecke, H., Schobert, B., Richter, H. T., Cartailler, J. P. & Lanyi, J. K. (1999). Journal of Molecular Biology 291, 899-911.

Lundstrom, K. (2005). Trends Biotechnol 23, 103-108.

Lusty, C. J. (1999). Journal of Applied Crystallography 32, 106-112.

MacAllister, R. J., Parry, H., Kimoto, M., Ogawa, T., Russell, R. J., Hodson, H., Whitley, G. S. J. & Vallance, P. (1996). British Journal of Pharmacology 119, 1533-1540.

Mackinnon, R. (2005). Science 307, 1425-1426.

McDevitt, C. A., Collins, R. F., Conway, M., Modok, S., Storm, J., Kerr, I. D., Ford, R. C. & Callaghan, R. (2006). Structure 14, 1623-1632.

Miroux, B. & Walker, J. E. (1996). J Mol Biol 260, 289-298.

Mitchell, D. A., Erwin, P. A., Michel, T. & Marletta, M. A. (2005). Biochemistry 44, 4636-4647.

Mitra, K., Schaffitzel, C., Shaikh, T., Tama, F., Jenni, S., Brooks, C. L., 3rd, Ban, N. & Frank, J. (2005). Nature 438, 318-324.

Moody, J. E., Millen, L., Binns, D., Hunt, J. F. & Thomas, P. J. (2002). Journal of Biological Chemistry 277, 21111-21114.

Murray-Rust, J., Leiper, J., McAllister, M., Phelan, J., Tilley, S., Santa Maria, J., Vallance, P. & McDonald, N. (2001). Nature Structural Biology 8, 679-683.

Murray, J. W., Rudino-Pinera, E., Owen, R. L., Grininger, M., Ravelli, R. B. & Garman, E. F. (2005). J Synchrotron Radiat 12, 268-275.

Navaza, J. (1994). Acta Crystallographica A 50, 157-163.

Niederweis, M. (2003). Molecular Microbiology 49, 1167-1177.

Notredame, C., Higgins, D. G. & Heringa, J. (2000). J Mol Biol 302, 205-217.

Ogawa, T., Kimoto, M. & Sasaoka, K. (1987). Biochemical & Biophysical Research Communications 148, 671-677.

Ogawa, T., Kimoto, M. & Sasaoka, K. (1989). Journal of Biological Chemistry 264, 10205-10209.

Orelle, C., Dalmas, O., Gros, P., Di Pietro, A. & Jault, J. M. (2003). Journal of Biological Chemistry 278, 47002-47008.

Orlowski, S., Selosse, M. A., Boudon, C., Micoud, C., Mir, L. M., Belehradek, J., Jr. & Garrigos, M. (1998). Cancer Biochem Biophys 16, 85-110.

Ostermeier, C. & Iwata, S. (2003).

Ostermeier, C., Iwata, S., Ludwig, B. & Michel, H. (1995). Nature Structural Biology 2, 842-846.

Ostermeier, C. & Michel, H. (1997). Current opinion in structural biology 7, 697-701.

Otwinoswki, Z. & Minor, W. (1997). Methods in Enzymology 276, 307-326.

Pannu, N. S. & Read, R. J. (1996). Acta Crystallogr A 52, 659-668.

Patzer, S. I. & Hantke, K. (1998). Molecular Microbiology 28, 1199-1210.

Pautsch, A. & Schulz, G. E. (1998). Nature Structural Biology 5, 1013-1017.

Perez-Moreno, M., Jamora, C. & Fuchs, E. (2003). Cell 112, 535-548.

Picot, D., Loll, P. J. & Garavito, R. M. (1994). Nature 367, 243-249.

Pinkett, H. W., Lee, A. T., Lum, P., Locher, K. P. & Rees, D. C. (2007). Science 315, 373-377.

Pos, K. M. & Dimroth, P. (1996). Biochemistry 35, 1018-1026.

Pos, K. M., Dimroth, P. & Bott, M. (1998). Journal of Bacteriology 180, 4160-4165.

Raina, S. & Missiakas, D. (1997). Annual Review of Microbiology 51, 179-202.

Ravelli, R. B. & Garman, E. F. (2006). Current opinion in structural biology 16, 624-629.

Ravelli, R. B. & McSweeney, S. M. (2000). Structure 8, 315-328.

Rosenberg, M. F., Callaghan, R., Modok, S., Higgins, C. F. & Ford, R. C. (2005). J Biol Chem 280, 2857-2862.

Rossig, L., Fichtlscherer, B., Breitschopf, K., Haendeler, J., Zeiher, A. M., Mulsch, A. & Dimmeler, S. (1999). Journal of Biological Chemistry 274, 6823-6826.

Röthlisberger, D., Pos, K. M. & Plückthun, A. (2004). FEBS Letters 564, 340-348.

Rubinfeld, B., Souza, B., Albert, I., Muller, O., Chamberlain, S. H., Masiarz, F. R., Munemitsu, S. & Polakis, P. (1993). Science 262, 1731-1734.

Sauna, Z. E., Nandigama, K. & Ambudkar, S. V. (2006). Journal of Biological Chemistry 281, 26501-26511.

Sauna, Z. E., Smith, M. M., Muller, M., Kerr, K. M. & Ambudkar, S. V. (2001). Journal of Bioenergetics & Biomembranes 33, 481-491.

Schneider, K., Dimroth, P. & Bott, M. (2000). Biochemistry 39, 9438-9450.

Schneider, T. R. & Sheldrick, G. M. (2002). Acta crystallographica 58, 1772-1779.

Schweizer, A., Roschitzki-Voser, H., Amstutz, P., Briand, C., Gulotti-Georgieva, M., Prenosil, E., Binz, H. K., Capitani, G., Baici, A., Pluckthun, A. & Grutter, M. G. (2007). Structure 15, 625-636.

Seddon, A. M., Curnow, P. & Booth, P. J. (2004). Biochimica et Biophysica Acta 1666, 105-117.

Sennhauser, G., Amstutz, P., Briand, C., Storchenegger, O. & Grutter, M. G. (2006). PLoS Biol 5, e7.

Sharma, S. & Davidson, A. L. (2000). Journal of Bacteriology 182, 6570-6576.

Shirai, H., Blundell, T. L. & Mizuguchi, K. (2001). Trends Biochem Sci 26, 465-468.

Smith, P. C., Karpowich, N., Millen, L., Moody, J. E., Rosen, J., Thomas, P. J. & Hunt, J. F. (2002). Molecular Cell 10, 139-149.

Sobczak, I. & Lolkema, J. S. (2003). Biochemistry 42, 9789-9796.

Sobczak, I. & Lolkema, J. S. (2004). Journal of Biological Chemistry 279, 31113-31120.

Sobczak, I. & Lolkema, J. S. (2005a). Biochemistry 44, 5461-5470.

Sobczak, I. & Lolkema, J. S. (2005b). Current Opinion in Microbiology 8, 161-167.

Sobczak, I. & Lolkema, J. S. (2005c). Microbiology & Molecular Biology Reviews 69, 665-695.

Stahlberg, H., Fotiadis, D., Scheuring, S., Remigy, H., Braun, T., Mitsuoka, K., Fujiyoshi, Y. & Engel, A. (2001). FEBS Letters 504, 166-172.

Stefkova, J., Poledne, R. & Hubacek, J. A. (2004). Physiological Research 53, 235-243.

Steuber, J., Krebs, W., Bott, M. & Dimroth, P. (1999). Journal of Bacteriology 181, 241-245.

Stieger, B., Fattinger, K., Madon, J., Kullak-Ublick, G. A. & Meier, P. J. (2000). Gastroenterology 118, 422-430.

Stieger, B. & Meier, P. J. (1998). Current Opinion in Cell Biology 10, 462-467.

Stirnimann, C. U., Grutter, M. G., Glockshuber, R. & Capitani, G. (2006). Cellular & Molecular Life Sciences 63, 1642-1648.

Thoma, R., Schulz-Gasch, T., D'Arcy, B., Benz, J., Aebi, J., Dehmlow, H., Hennig, M., Stihle, M. & Ruf, A. (2004). Nature 432, 118-122.

Tong, L. & Rossmann, M. G. (1997). Methods Enzymol 276, 594-611.

Tran, C. T., Leiper, J. M. & Vallance, P. (2003). Atherosclerosis Supplements 4, 33-40.

Tran, C. T. L., Fox, M. F., Vallance, P. & Leiper, J. M. (2000). Genomics 68, 101-105.

Tusnady, G. E. & Simon, I. (2001). Bioinformatics 17, 849-850.

Vallance, P. & Chan, N. (2001). Heart 85, 342-350.

Vallance, P. & Leiper, J. (2002). Nature Reviews. Drug Discovery 1, 939-950.

Van den Berg, B., Clemons, W. M., Jr., Collinson, I., Modis, Y., Hartmann, E., Harrison, S. C. & Rapoport, T. A. (2004). Nature 427, 36-44.

van der Does, C. & Tampe, R. (2004). Biological Chemistry 385, 927-933.

Van der Rest, M. E., Abee, T., Molenaar, D. & Konings, W. N. (1991). European Journal of Biochemistry 195, 71-77.

van der Rest, M. E., Molenaar, D. & Konings, W. N. (1992). Journal of Bacteriology 174, 4893-4898.

van Es, J. H., Barker, N. & Clevers, H. (2003). Current Opinion in Genetics & Development 13, 28-33.

van Geest, M. & Lolkema, J. S. (1999). Journal of Biological Chemistry 274, 29705-29711.

van Geest, M. & Lolkema, J. S. (2000). Biochimica et Biophysica Acta 1466, 328-338.

van Veen, H. W., Margolles, A., Muller, M., Higgins, C. F. & Konings, W. N. (2000). EMBO Journal 19, 2503-2514.

Vandeputte-Rutten, L., Kramer, R. A., Kroon, J., Dekker, N., Egmond, M. R. & Gros, P. (2001). EMBO Journal 20, 5033-5039.

Wagner, S., Bader, M. L., Drew, D. & de Gier, J. W. (2006). Trends in Biotechnology 24, 364-371.

Wilson, A. (1949). Acta crystallographica 2, 318-321.

Yernool, D., Boudker, O., Jin, Y. & Gouaux, E. (2004). Nature 431, 811-818.

Yuan, Y. R., Blecker, S., Martsinkevich, O., Millen, L., Thomas, P. J. & Hunt, J. F. (2001). J Biol Chem 276, 32313-32321.

Zhang, H., Kurisu, G., Smith, J. L. & Cramer, W. A. (2003). Proc Natl Acad Sci U S A 100, 5160-5163.

Chapter H

Acknowledgement

This PhD could not have been performed without the support, contribution and understanding of numerous people that were and still are important to me during the last years.

First of all, I want to thank my thesis supervisor Prof. Dr. Markus G. Grütter who gave me the opportunity to work in his group and made it possible for me to work on the interesting and challenging topic of membrane proteins. He gave me a lot of freedom during my work and his helpful ideas and permanent support were invaluable to the progress of the projects.

I am indebted to Prof. Milan Vašák for having accepted to be co-referee of my thesis and I am grateful for the helpful and fruitful discussions I could have with him on the DDAH project.

I want to thank Prof. Dr. Raimund Dutzler for his continuous comments and suggestions related to the work on membrane proteins and especially, to have accepted to be one of the co-referees of my thesis.

I want to thank Prof. Dr. Andreas Plückthun for the discussions and comments related to the project CitS and the fab f3p4 structure.

I thank Prof. Dr. Bruno Stieger and Dr. Daniel Noe for their help and scientific input related to the work on mBsep.

Many thanks go to Oliver Braun and Markus Knipp for the successful collaboration on the protein DDAH-1.
I thank Thomas Huber for the collaboration on the project CitS. We had valuable and fruitful discussions during all stages of the project, and his great enthusiasm and dedication to the project were very motivating. In addition, I thank Daniela Röthlisberger for generation of the CitS binding Fab fragments.

I am thankful to Prof. Dr. Konrad Basler of the University of Zurich and his co-worker Dr. Sandipan Chaterijee for the cooperation and discussions within the β-catenin project.

In special I want to thank Annette Haisch and Beat Blattmann for their humor, questions, comments and enthusiasm during the work. I thank Daniela Deriu, Lucas Lüthy and Markus Kaufmann who were helping me a lot during the early stages of my PhD. I am also thankful to Susan Firbank for teaching me the essentials in crystallography and for the correction of my thesis.

Furthermore I want to thank all present and former members of the Grütter, Plückthun and Dutzler groups and the "Büro". They all together provided a nice and fruitful atmosphere in and also outside of the lab with a special thanks to the "Endstation Jass" members.
Last but definitely not least I want to thank my family for the continuous support during my graduate studies and my PhD.

Curriculum vitae

Personal Data

Name	Frey
First Name	Daniel
Date of Birth	17th of August 1973
Place of Birth	Basel Switzerland
Nationality	German

Education:

1984 – 1993	Gymnasium, Hans-Thoma Gymnasium Lörrach, Germany
1993	Abitur
1994 – 2000	Undergraduate Studies, Albert-Ludwigs-Universität Freiburg i. Br., Germany
1996	Vordiplom, Albert-Ludwigs-Universität Freiburg i. Br., Germany
2000	Diploma thesis at the Institute of Organic Chemistry and Biochemistry, Albert-Ludwigs-Universität Freiburg i. Br., Germany. Subject: **Reinigung, Charakterisierung und Kristallisation des Porin A aus Mycobacterium Smegmatis**, supervisor Prof. G.E. Schulz.
2000 – 2007	Graduate studies at the Department of Biochemistry, University of Zürich, supervisor Prof. M.G. Grütter.
2002	Practical course **Modern Protein Technologies** of the NCCR in Structural Biology
2004	Practical course **Membrane Protein Expression, Purification, 2D Crystallization and Imaging** of the NCCR in Structural Biology
2005	Practical course **Synchrotron Data Acquisition Techniques in Macromolecular Crystallography Technologies** of the NCCR in Structural Biology
2007	Dissertation **Biochemical and Structural Studies of Dimethylarginine Dimethylaminohydrolase-1, the ABC Transporter superfamily and the Sodium Citrate transporter CitS.**

Publications and Meeting contributions

Publications:

Daniel Frey, Oliver Braun, Christophe Briand, Milan Vašák and Markus G. Grütter.
Structure of the Mammalian NOS Regulator Dimethylarginine Dimethylaminohydrolase (DDAH-1): A Basis for the Design of Specific Inhibitors.

Daniel Frey, Thomas Huber, Andreas Plückthun and Markus G. Grütter.
Unusual twinning in the crystal structure of the Fab f3p4 (in preparation)

Thomas Huber, Daniel Frey, Andreas Bosshart, Markus G. Grütter and Andreas Plückthun.
Structures of CitS binding Ankyrins (in preparation)

Meeting contributions:

Daniela Deriu, Daniel Frey and Markus Grütter. **Towards the structure of $GABA_B$ receptor.** *1st International Conference on Structure, Dynamics and Function of Proteins in Biological Membranes, 2001 Monte Verità, Ascona, Switzerland.*

Daniel Frey, Annette Haisch, Susan Firbank & Markus Grütter. **Strategy for Structural Investigation of ABC Transporters.** *NCCR symposium, 2002, Zurich, Switzerland.*

Daniel Frey, Christophe Briand, Gaby Sennhauser, Michaela Kramer, Annette Haisch & Markus Grütter. **Expression, purification and characterization of Bacterial ABC transporters.** *BioMedical Transporters 2005, St-Gallen, Switzerland.*

Thomas Huber, Daniela Röthlisberger and Andreas Plückthun
Daniel Frey and Markus Grütter
Binding Partners for Co-Crystallization of the Na^+-Citrate Symporter CitS NCCR. *3rd International Conference on Structure, Dynamics and Function of Proteins in Biological Membranes, 2006 Monte Verità, Ascona, Switzerland.*

Thomas Huber, Daniel Steiner, Daniela Röthlisberger and Andreas Plückthun
Daniel Frey and Markus Grütter
Binding Partners for Co-Crystallization of the Na^+-Citrate Symporter CitS
NCCR symposium, 2006, Zurich, Switzerland.